Selection and characterization of DNA aptamers

Vincent J.B. Ruigrok

Selection and characterization of DNA aptamers

Vincent J.B. Ruigrok

Thesis committee

Promotors
Prof. Dr. J. van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University

Prof. Dr. H. Smidt
Personal chair at the Laboratory of Microbiology
Wageningen University

Other members
Prof. Dr. W.J.H. van Berkel, Wageningen University
Dr. M.H.M. Eppink, Synthon BV, Nijmegen
Dr. W. Haasnoot, RIKILT-Institute of Food Safety, Wageningen
Dr. J.H.G. Lebbink, Erasmus University Medical Center, Rotterdam

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Selection and characterization of DNA aptamers

Vincent J.B. Ruigrok

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 21 June 2013
at 4 p.m. in the Aula.

Vincent J.B. Ruigrok
Selection and characterization of DNA aptamers
160 pages.
PhD thesis, Wageningen University, Wageningen, NL (2013)
With references, with summaries in Dutch and English
ISBN 978-94-6173-564-5

The key is man's power of accumulative selection: nature gives successive variations; man adds them up in certain directions useful to him.

- Charles Darwin -

Evolution is not a law, it is just a phenomenon.

- Jacques Monod -

Table of contents

Note: this digital edition contains hyperlinks for fast navigation. Click on a chapter name to navigate to it, click on a page number to return tot the table of contents. Return to the start of a chapter by clicking the page header. Original publications can be accessed on the publishers website from the "list of publications", or from the chapters title page.

Preface and outline

Preface and outline

Nucleic acids are versatile molecules, but that might not seem obvious at first glance. DNA is primarily known for its double helical appearance, and as carrier of the blueprint of life; RNA is mainly known as essential intermediate in translating the genetic code to proteins. Both DNA and RNA, however, serve other purposes that can even be unrelated to their genetic content. In 1990, researchers selected specific RNA molecules that were capable of binding molecules that normally do not interact with RNA. These molecules were selected from a large pool of RNA molecules, by a process that was named systematic evolution of ligands by exponential enrichment, or SELEX. The resulting RNA molecules were named 'aptamers' which roughly means 'fitting part'. Two years later, a single stranded DNA molecule was identified that specifically binds thrombin, a protein that normally does not interact with DNA. This served as the first example of a DNA aptamer. In the following years, numerous DNA and RNA aptamers, capable of binding a wide variety of targets, became available, leading to a plethora of potential applications. The capability of binding a specified target results from an intricate interplay between the nucleotide composition of the DNA or RNA molecule, the three-dimensional shape that this composition enables, and specific features of the target itself. Unfortunately, no accurate predictions can be made on which nucleotide composition will allow an aptamer to bind a specified target, therefore aptamer selection remains largely a trial and error approach. Selection of DNA aptamers and subsequent characterization of the aptamer-target interaction is the main focus of this thesis.

Besides aptamers, various other groups of molecules are capable of binding specified target molecules. Each group has its own advantages and disadvantages, when compared to the others. In **Chapter 1** four of such groups are described in general terms and their specific features are compared.

One advanced technique to study aptamer-target interactions is Surface Plasmon Resonance (SPR). With this technique, binding interactions can be monitored in real-time, allowing for gaining valuable insights in the speed of complex formation and dissociation. **Chapter 2** provides an overview on how to initiate SPR experiments and how to correctly process and analyze kinetic data.

In **Chapter 3** the selection and subsequent characterization of streptavidin-binding aptamers is described. In this work five aptamers were identified and their binding kinetics were determined using SPR. Furthermore, the aptamers and the aptamer-streptavidin complex were studied by native mass spectrometry and small angle X-ray scattering.

In order to gain additional information on the aptamer-streptavidin complex, crystallization trials were performed in an attempt to co-crystallize the complex. These trials, and other methods for the characterization of the molecular basis of aptamer-protein interactions, are described in **Chapter 4**.

SpaC, a subunit of pili present on the probiotic Gram-positive bacterium *Lactobacillus rhamnosus* GG, was also used as target for aptamer selection. Cloning of the gene and recombinant production of SpaC protein are described in **Chapter 5**, as well as the characterization of enriched DNA oligonucleotides.

Experimental conditions for aptamer selection are often based on practical considerations and convenience, partly because the dynamics of aptamer selection are not fully understood. To gain a more fundamental insight in the dynamics of selection, samples of multiple selection rounds, obtained during distinct experiments, were barcoded, pooled and subjected to a multiplexed high throughput sequencing approach. Results and practical details of this approach are described in **Chapter 6**.

A new application of SPR is presented in **Chapter 7**: a method for capturing supercoiled plasmid DNA. The method makes use of a triple helix forming oligonucleotide and facilitates studying protein-DNA interactions on DNA with a physiologically relevant topology.

Finally, **Chapter 8** summarizes the research described in this thesis and discusses new developments that are beneficial for aptamer selection. Future opportunities for aptamer applications are discussed as well.

Chapter 1

Alternative affinity tools: more attractive than antibodies?

Vincent J. B. Ruigrok[†], Mark Levisson[†] , Michel H. M. Eppink, Hauke Smidt and John van der Oost

[†] Contributed equally

Biochemical Journal, 2011. 436(1): p. 1-13

Abstract

Antibodies are the most successful affinity tools used today, in both fundamental and applied research (diagnostics, purification and therapeutics). Nonetheless, antibodies do have their limitations, including high production costs and low stability. Alternative affinity tools based on nucleic acids (aptamers), polypeptides (engineered binding proteins) and inorganic matrices (molecular imprinted polymers) have received considerable attention. A major advantage of these alternatives concerns the efficient (microbial) production and *in vitro* selection procedures. The latter approach allows for the high-throughput optimization of aptamers and engineered binding proteins, e.g. aiming at enhanced chemical and physical stability. This has resulted in a rapid development of the fields of nucleic acid- and protein-based affinity tools and, although they are certainly not as widely used as antibodies, the number of their applications has steadily increased in recent years. In this chapter, we compare the properties of the more conventional antibodies with these innovative affinity tools. Recent advances of affinity tool developments are described, both in a medical setting (e.g. diagnostics, therapeutics and drug delivery) and in several niche areas for which antibodies appear to be less attractive. Furthermore, an outlook is provided on anticipated future developments.

Introduction

Many of the current applications in life sciences and biotechnology require the use of highly selective binders for the detection, purification or removal of specific molecules. The tools that allow these processes to proceed are generally referred to as affinity tools. At present, antibodies of mammals are the best characterized and most widely used affinity tools. Natural antibody targets (antigens) are surface molecules of invading entities (bacteria and viruses), including (poly)peptides and carbohydrates. However, owing to their unlimited variability, an infinite number of organic molecules are potential antigens that can be captured with very high affinity and specificity. Nowadays, antibodies can easily be produced using a range of techniques by either the immune system or synthetic libraries [101]. Nevertheless, antibodies do have their limitations [16]. They are reasonably sensitive to harsh conditions and are rapidly inactivated under acidic conditions, in the presence of proteases or at elevated temperatures. In addition, antibodies are large multi domain protein complexes with specific disulfide bonds and glycosylations. This implies that producing antibodies is generally difficult and expensive. Moreover, the extensive use of antibodies and their applications (e.g. as therapeutic agents) has resulted in a complicated patent situation [143]. Therefore, in recent years, other affinity tools, such as engineered binding proteins, aptamers and molecular imprinted polymers (MIPs), have gained a lot of interest as potential alternatives for antibodies in many applications. An overview of selected companies that are currently developing new affinity tools as alternative to antibodies is provided in Supplementary Table S1.1 (at page 28 and http://www.BiochemJ.org/bj/436/bj4360001add.htm).

The new affinity tools have several advantages compared with antibodies. Attractive features include a reduced molecular mass, an enhanced stability, a more efficient selection and screening procedure, and cost-effective production methods. In this chapter, we aim to provide an overview of the current state of antibodies, engineered binding proteins, aptamers and MIPs, as affinity tools for different applications. We compare the properties and production methods of each tool and provide information on screening and binder-selection procedures. Furthermore, we elaborate on the potential of each affinity tool and relate this to various applications in a medical setting (diagnostics, therapeutics and drug delivery) and several niche applications for which antibodies appear to be less suitable.

Antibodies

Antibodies recognize their target with high specificity, and they generally bind it with a very high affinity, typically in the nano/pico-molar range. Depending on their origin, antibodies are polyclonal or monoclonal. A polyclonal antibody (pAb) is purified from blood serum of immunized mammals; it is a set of antibody variants that bind different epitopes of a target antigen, and are therefore not very specific.

A monoclonal antibody (mAb) is a single antibody variant that is derived from a single cell line (hybridoma) [139]; mAbs bind only one epitope on a single antigen, making them more specific than pAbs [37].

Because of their high specificity and affinity, antibodies are useful for a wide variety of applications. Antibodies, especially mAbs, are nowadays the most widely used tool for diagnostic applications in fundamental and applied research. Furthermore, because of their potential to specifically interact with (neutralize) a target molecule, several U.S. Food and Drug Administration (FDA)-approved mAbs are used in the clinic as therapeutics in the treatment of diseases ranging from inflammation to cancer [195, 224].

Hybridoma: an endless antibody supply

Once an antibody-producing hybridoma cell line is established, it provides an endless supply of mAbs, because hybridomas are fusions of antibody-producing spleen cells and immortal myeloma cells [139] (Figure 1.1). Immunization of mice with antigen is the first step to establish a hybridoma. An appropriate immunization protocol is critical for provoking an adequate immune response. Several strategies have been developed to increase throughput and success rate of immunization and to decrease the number of animals needed, including single-step immunization [319], genetic immunization [32] and multiplex immunization [198]. Only when an immune response is confirmed, can mouse spleen cells be harvested and fused with myeloma cells to continue hybridoma selection. After cell fusion, the mixture of hybridomas is generally diluted in such a way that each resulting micro culture is derived from a single cell. This limiting dilution cloning (LDC) procedure is simple, but it is time-consuming and has a low throughput. Several other selection methods with higher throughput, based on flow cytometry and cell sorting, have been established [23]. The next step is screening of supernatants, to identify hybridomas that produce antibodies with the desired properties. To this end, the antigen microarray assay, basically a miniaturized ELISA, reduces the amount of time and reagents required compared with conventional ELISA screening [50]. Once appropriate

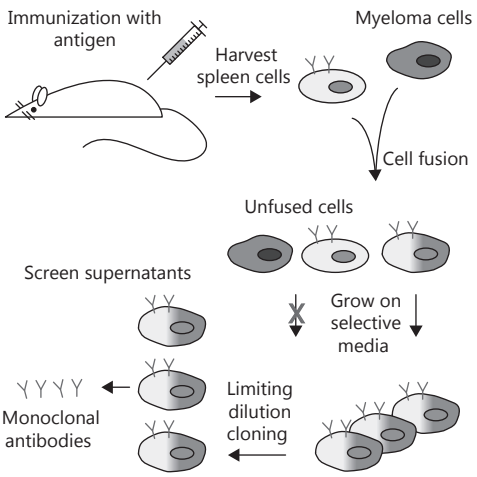


Figure 1.1: Hybridoma selection. When a mouse is injected with a certain antigen and detection of an immune response is confirmed, spleen cells are harvested and fused with myeloma cells. Fusion products are grown in selective media to select for fused cells. After LDC of fused cells, cell supernatants are screened for mAbs with the desired antigen specificity.

hybridomas are selected, they can be preserved and used for antibody production.

Regardless of whether antibodies are polyclonal or monoclonal, the most commonly used antibodies are of the IgG type. IgGs are Y-shaped and consist of two longer 'heavy' chains and two shorter 'light' chains bound together by disulfide bonds (Figure 1.2A). The Fab region consists of constant and variable domains of the heavy and light chain. Antigen binding takes place at the complementarity-determining regions (CDRs), which are located in the variable domains (V_H and V_L) in the Fab region. The Fc region consists of the constant heavy chain domains; it maintains stability and is involved in other interactions of the immune system.

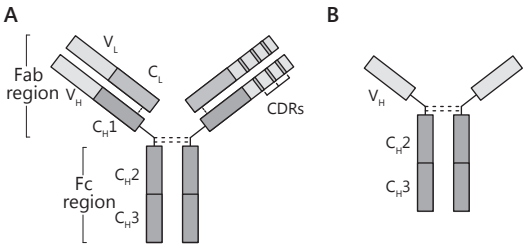


Figure 1.2: Schematic structures of a human IgG and heavy chain antibody. A) Human IgG. Two heavy and light chains are linked together by disulfide bonds. The Fab region consists of constant (C) and variable (V) domains of the heavy (H) and light (L) chain. CDRs are located in the variable domains (V_H and V_L) in the Fab region. The Fc region consists of the constant heavy chain domains. B) Heavy chain antibody. Two heavy chains, consisting of two constant regions C) and one variable region (V), are linked together by disulfide bonds.

Medical applications

Antibodies are often used in biochemical laboratories for numerous routine diagnostic tests such as ELISA and Western blotting. In a medical setting, these tests are generally used for the detection of infectious and parasitic diseases. In addition, ELISA is also often used for routine screening of food and environmental samples for microbial and chemical contamination. It is difficult to multiplex ELISAs because of cross-reactivity of secondary antibodies, which are required for enhanced sensitivity [322]. The pregnancy test is a well-known example of a cheap and easy to use diagnostic test that relies on antibodies [detection of the pregnancy-associated hormone hCG (human chorionic gonadotropin)]. Although this pregnancy test has been on the market for over 35 years [26], development of similar tests for other targets has been slow. However, increased interest in point-of-care diagnostics has rapidly increased the number of diagnostic tests available within the last 10 years [293], e.g. in 2004, the FDA approved the first rapid test for HIV in oral fluids [20]. The number of point-of-care applications is expected to increase steadily in the near future.

One of the most advanced applications of antibodies to date, however, is their use as therapeutic agents. Muromonab-CD3 (Orthoclone OKT3), an anti-CD3 antibody, became the first murine mAb to be approved by the FDA in 1986 as a therapeutic agent to prevent organ rejection after transplantation [79]. Administration of these murine antibodies, however, carried the risk of an undesired immune response

(immunogenicity) and other side effects [296]. It took until 1994 before the second mAb [Abciximab (ReoPro), a chimaeric anti-GPIIb/IIIa antibody] received FDA approval, partly because the side effects of these murine mAbs had to be reduced; to date, over 30 mAbs are FDA-approved [9]. Technological advances have enabled production of chimeric and later humanized antibodies (a human IgG scaffold with only the CDRs of murine origin), which have fewer side effects than their murine counterparts; the current state-of-the-art is a fully human antibody, produced by transgenic mice [157].

Therapeutic mAbs can operate through different mechanisms. Some mAbs act independently of the Fc domain of the antibody by blocking the interaction between a receptor and the receptor molecule (by binding to either one of them), or by inducing a signal transduction cascade upon receptor binding. Other mechanisms do require the Fc domain to recruit components of the immune system to lyse target cells by either antibody-dependent cell-mediated cytotoxicity or complement-dependent cytotoxicity [161].

Antibodies are also used for targeted drug delivery. For this purpose, mAbs are labelled with radioisotopes or toxic drugs, combining antibody specificity with enhanced toxicity because of their label. Two radioimmunoconjugates [Ibritumomab tiuxetan (Zevalin), anti-CD20, and Tositumomab-131I (Bexxar), anti-CD20] are FDA-approved [310], whereas the only FDA-approved conjugate with a toxin [Gemtuzumab ozogamicin (Mylotarg), anti-CD33] has recently been withdrawn from the market because of concerns about the product's safety and clinical benefits to patients [61]. Radioimmunoconjugates can also be used for *in vivo* imaging of tumors, to evaluate targeting or success of therapy. Specific binding of antibodies to their target allows for a very good signal-to-noise ratio. This imaging is Fc-independent, therefore conjugates of antibody fragments could be used for imaging as well.

Drawbacks of antibodies

High selectivity and affinity make antibodies especially suitable for diagnostic and therapeutic applications. Antibodies have a large and complex structure, which allows them to bind antigens and simultaneously recruit components of the immune system and, as we have seen before, this is an essential property in some forms of immunotherapy. However, owing to their complex structure, antibodies are susceptible to degradation, aggregation, modification (e.g. oxidation or deamidation) and denaturation. Moreover, the bulky nature of antibodies also limits their potential for some other applications; in affinity chromatography, smaller binders are preferred because of enhanced efficiency, whereas smaller binders exhibit an improved signal-to-noise ratio when applied to molecular imaging.

Antibodies have not yet been successfully produced in simple microbial hosts, because of their large complex structure and the required specific post-translational glycosylation. Hence they have to be produced in mammalian cell lines instead

[e.g. NS0, Per.C6®, Chinese-hamster ovary (CHO) and human embryonic kidney (HEK)-293 cells], the cultivation of which is complex and costly (expensive media, long fermentation lead times, scaling issues, use of gases). Although nowadays mAbs are well produced and purified under mild conditions, they nonetheless remain sensitive to aggregation, deamination and oxidation. The use of antibodies is therefore often restricted to conditions resembling their physiological environment. In the presence of organic solvents or in other non-physiological conditions (e.g. elevated temperature, high pH and high salt concentration), they generally lose their function.

When administered intravenously, antibodies can have half-lives of several weeks [29]. A long half-life is desirable for some applications, because it increases efficacy and reduces the required dose. On the other hand, for certain applications, short half-lives are preferred, for instance when rapid clearance of toxic immunoconjugates is required to reduce whole-body exposure.

Size-related limitations of antibodies can be alleviated by using specific parts of antibodies that can be produced in microbial hosts (e.g. Fab regions or variable domains). These parts are smaller and, as such, they can reach deeper into tumors; in addition, half-lives can be tailored [29]. Removing fragments of the antibody that are not directly involved in target binding, however, can negatively affect affinity. In addition, antibody fragments tend to aggregate, making it difficult to purify them [95].

An alternative for antibody fragments are heavy chain antibodies (Figure 1.2B). Heavy chain antibodies are natural variants that lack a light chain; they are isolated from camelid species [86] and cartilaginous fish [266]. Heavy chain antibodies from camelid species are much smaller than classic antibodies, and, like artificially truncated antibody fragments, they can be produced in microbial hosts. As such, they are being optimized for certain specificities [87, 299, 318], by using high-throughput analyses such as phage display (see below).

Engineered binding proteins

In parallel to the development of antibody-based binding proteins, such as the antibody fragments and single-domain antibodies, independent studies have focused on modifying non-antibody proteins into binding proteins, with the goal of developing binding proteins with improved features [9]. The concept of engineering new binding functions is based on the molecular composition of antibodies (Figure 1.2). The typical structural element of an antibody is a well-conserved rigid scaffold on which highly variable loops (CDRs) are positioned. Similarly, binding proteins can be isolated from large libraries of protein variants with a constant framework and randomized variable regions that can, in principle, interact with any target molecule. These new binding proteins have great potential as affinity tools in various biotechnological applications and as therapeutic agents.

The generation of novel binding proteins

An important aspect of successful design and engineering of a binding protein scaffold is the ability to generate and analyze a large number of mutated derivatives. A powerful high-throughput technology for evolution-driven engineering is molecular display: the generation of large (poly)peptide libraries and subsequent selection for variants with desired biological and physicochemical properties. Display technologies are based on a physical link between a protein and its encoding gene, hence coupling phenotype and genotype. The most commonly used display technology is phage display [259]. However, other methods such as bacterial and yeast cell-surface display [46 67], ribosome display [156, 320] and mRNA display [156] have also been successfully applied. In this chapter, we will restrict ourselves to describing phage display for the selection of engineered binding proteins.

In phage display technology, nucleotide sequences encoding variants of peptides, antibodies or proteins are fused to a gene that encodes a phage coat protein. After correct assembly, phage particles display the encoded (poly)peptide on their surface [217]. The most widely used vectors for library construction are based on the filamentous phages fd, M13 and related phagemids [133, 213, 258]. However, display systems based on other phages and viruses have also been developed, including the lytic phages λ and T7 [21]. A current shortcoming of phage display is that only a limited number of vectors, libraries and complete systems are commercially available.

In general, when a phage display procedure with an M13 phagemid vector is used, a library of variant DNA sequences encoding a protein of interest is created using *in vitro* evolution techniques [181]. These variants are subsequently cloned into the phagemid vector as fusion to a coat protein gene (frequently the pIII-gene) (Figure 1.3). *Escherichia coli* cells are transformed with the phagemids and then infected with a helper-phage, generating a library of phages displaying the variant proteins. The phage library is exposed to an immobilized target molecule, and the phages with appropriate specificity and affinity are captured. The non-binding phages are washed off, although some non-specific binding can occur. Bound phages are eluted by conditions that disrupt the interaction between the displayed protein and the target. Eluted phages are then used to re-infect *E. coli* cells. The resulting amplified phage population is a secondary library that is highly enriched in phages displaying proteins that bind to the target. In an iterative process, these steps are repeated using washing steps with increasing stringency (bio-panning), resulting in a phage population enriched in a limited number of variants with the desired binding affinity and specificity. After several rounds of bio-panning (generally three to five), monoclonal phages may be selected and analyzed individually. Target-binding (poly) peptides are identified by DNA sequence analysis of the phage and may subsequently be used as input material for another series of *in vitro* evolution in order to obtain a (poly)peptide with even higher binding affinity.

The diversity of protein scaffolds

In the last two decades, over 50 new alternative non-Ig protein scaffolds have been reported as potential affinity tools. These protein scaffolds comprise an extremely diverse group of binding molecules, which differ in many aspects such as origin, size, structural topology, engineering strategies, mode of interaction and applicability. The classification of these protein scaffolds is most often based on their structures and the strategies applied for engineering the binding affinity. Two examples of protein scaffolds that have been used successfully are discussed in some more detail below; these and other protein scaffolds have been the focus of several excellent reviews [16, 74, 83, 90, 105, 203, 204, 216, 260].

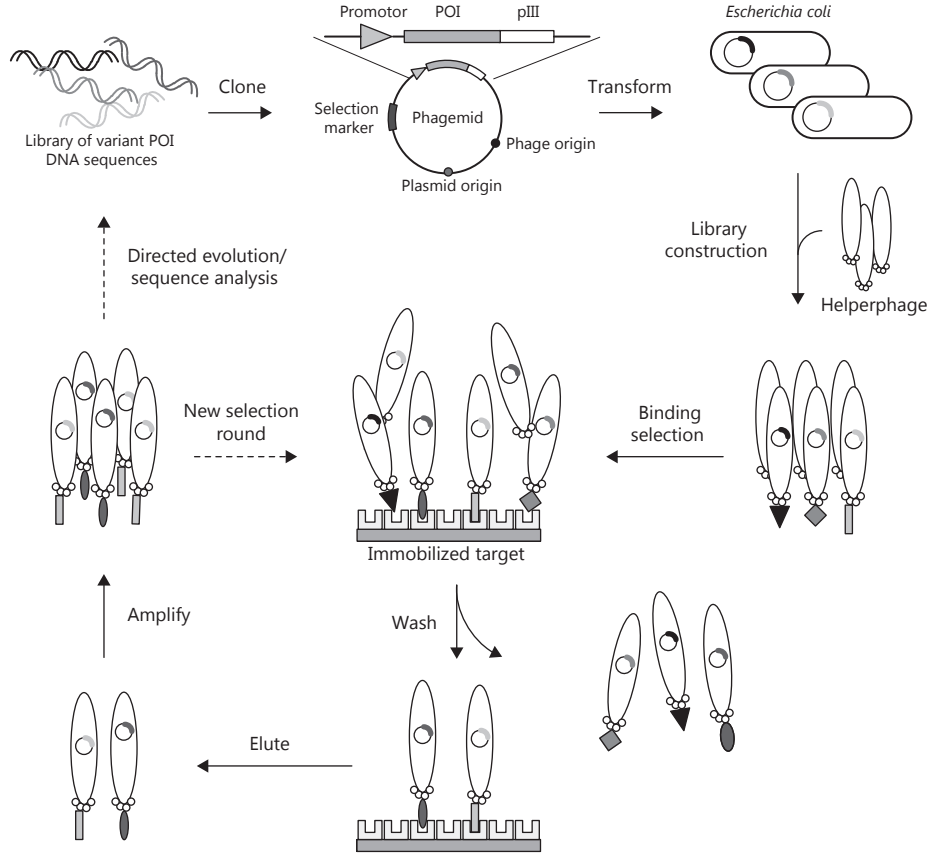


Figure 1.3: M13 phage display cycle. A library of DNA fragments encoding random variants of the protein of interest (POI) is created (fused to gene encoding M13 coat protein pIII) and cloned into a phagemid vector. *Escherichia coli* cells are transformed with the obtained constructs and subsequently infected with helper-phage to create a library of phages, each displaying a variant of the protein of interest. The library is exposed to an immobilized target molecule. Non-binding phages are washed away. Bound phages are eluted and then amplified by infecting bacterial cells. This selection and amplification process can be repeated as necessary using more stringent washing conditions in order to obtain phages with the displayed protein of interest variants with the highest target-binding affinity. Finally, the DNA of high-affinity binding phages can be sequenced or subjected to another round of evolution.

Fibronectin type III domain (FN3) belongs to the immunoglobulin-like β-sandwich class of scaffold proteins. FN3 is a small 10 kDa domain occurring in many animal proteins involved in ligand binding (e.g. cell-surface receptors), and is one of the most widely used scaffolds today [17, 140]. The β-sandwich structure consists of seven β-strands with three connecting loops on one end of the sheet, closely resembling the structure of an Ig variable domain (Figure 1.4A). Initially, two of the surface loops were randomized, and binders were selected using phage display [140]. In a more recent study, all three surface loops were randomly varied, after which the best binders were successfully selected using mRNA display [312]. At AdNexus Therapeutics, FN3-based scaffold proteins (AdNectins) are being used to develop novel cancer therapeutics (see Table S1.1). The first AdNectin product concerns a protein designed to block the VEGFR2 [vascular endothelial growth factor (VEGF) receptor 2] signal transduction pathway, inhibiting the growth of new blood vessels in tumors.

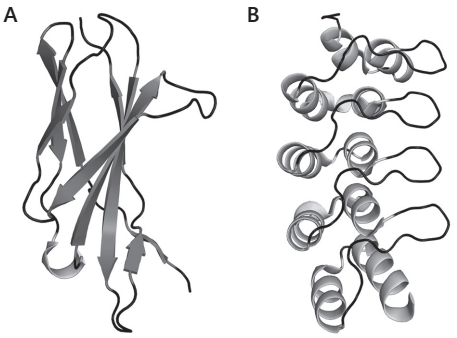


Figure 1.4: Three-dimensional structures of engineered binding proteins. The overall structures of A) an FN3 monobody (PDB code: 3K2M [309]) and B) a DARPin (PDB code: 1SVX [15]) are shown.

Designed ankyrin repeat proteins (DARPins) belong to the repeat protein class of scaffold proteins. DARPins are small (14–21 kDa) modular proteins derived from natural ankyrin repeat proteins, which are versatile binding proteins, the corresponding genes of which are highly abundant in the human genome [271, 272]. DARPins have a modular architecture consisting of a single structural motif that occurs several times, and, after stacking, they form an elongated repetitive domain that is shielded by two capping repeats (Figure 1.4B). Libraries have been constructed in which residues residing in a loop and on a helix were randomized in each repeat; subsequently multiple repeats were randomly assembled [15]. The selection method most frequently used with DARPin libraries is ribosome display. High-affinity binders have been selected against targets such as maltose-binding protein [15] and an anti-IgE antibody [292].

Considerations for choosing the right scaffold

The wide variety of potential applications, such as therapeutics, diagnostics, affinity purification and molecular imaging, has led to the exploration of many different protein scaffolds. However, depending on the intended application, potential selection criteria include origin, size and structure of the protein to be used as a scaffold. For eventual commercial application of a selected scaffold protein, desirable features are their efficient production, high solubility, stability, specificity and a favorable intellectual property (IP) situation.

It is very important that the selected protein has a compact and structurally rigid core framework that has a relatively high intrinsic stability. This is an obvious requirement to generate a library by randomization of solvent-exposed amino acid residues, without affecting the scaffold's fold and half-life. The interface topography is a major determinant of protein-protein interactions. Therefore, if protein binding is intended, the shape of the selected protein's binding site (convex, flat or concave) could influence epitope selection and the binding mode [216].

In case the binding proteins are intended for therapeutic use, a potential problem of novel scaffolds may be their immunogenicity. Therefore, considerations should be made about the origin of the protein used as a scaffold; compared with a protein from a distinct origin, a human protein is less likely to cause an undesired immune response. Another disadvantage of using therapeutic proteins of non-human origin is the fact that they may face additional problems from regulatory authorities with respect to immunogenicity issues. However, it should be mentioned that human scaffold proteins may become immunogenic when multiple surface-exposed amino acids are randomized or when their binding site is altered.

Small size is often a desired feature of the selected protein used as a scaffold. For instance, in molecular imaging of tumor cells, the binding protein should rapidly find its target molecule in the patient, whereas unbound binding proteins should be quickly excreted; this will result in high-contrast tumor imaging and a reduced time between injection and imaging. Because small binding proteins tend to require less time for target recognition or for excretion, bigger scaffold proteins (e.g. antibodies) are less attractive for this particular application.

Stability of the selected protein scaffold is of great importance when used for affinity-purification applications. The protein should be sufficiently resistant to elevated temperatures, organic solvents, detergents and pH changes; in addition, the column material should preferably be cleanable and reusable. Most of the above-described conditions are detrimental to proteins, especially antibodies, and it might therefore be interesting to investigate proteins from thermophilic micro-organisms as candidate scaffolds. Proteins from these thermophiles generally display a high intrinsic thermal and chemical stability [80, 289].

The above considerations indicate that non-antibody protein scaffolds do not constitute a homogeneous group with similar properties. It is difficult to predict which of these diverse proteins are most suitable for a certain application. Although several of these binding proteins may have a lot of potential for biotechnological applications, the development of these alternatives to antibodies is still in an early phase of proof-of-principle evaluation. With respect to medical applications, only a few have progressed to clinical trials. Furthermore, expertise is often limited to single laboratories, and only a few libraries of protein variants are commercially available [282].

Aptamers

Aptamers are oligonucleotide ligands that are selected for high affinity binding to molecular targets. Both RNA- and single-stranded DNA (ssDNA)-based aptamers have been described, generally with a size of 15–60 nt. Reported aptamer targets range from small organic molecules such as ethanolamine and acetylcholine [24, 165] to large protein complexes [18, 285] and even cells [27, 281]. The term 'aptamer' is derived from the Latin word 'aptus' (to fit) and the Greek word 'meros' (part or piece). Aptamers can have affinities in the nanomolar range, which is comparable with that of mAbs [118]. Target-binding aptamers are generally selected using a so-called systematic evolution of ligands by exponential enrichment (SELEX) procedure, reported independently by two groups in 1990 [58, 285].

The choice for either DNA or RNA as a basis for aptamer development depends on practical considerations and partly on the final application. RNA has a relatively flexible backbone compared with DNA, and as such it has a broader range of potential target molecules. However, an obvious practical drawback of RNA is the fact that it is more prone to chemical and enzymatic degradation. In addition, selection of RNA aptamers is more laborious as its processing requires more enzymatic steps. Although RNA stability issues can partly be overcome by using modified nucleotides (e.g. spiegelmers, see below), it might be desirable for some applications (e.g. biosensors or affinity chromatography) to use the naturally more robust DNA aptamers. To date, however, the only FDA-approved therapeutic aptamer (Pegaptanib, see below), is an RNA aptamer.

Aptamers are selected in vitro

In contrast with the classical production of antibodies in animals, all steps of a SELEX procedure take place in a test tube. The *in vitro* SELEX procedure consists of multiple selection rounds, enriching proper binders in each subsequent round. The principle of iterative enrichment of binding oligonucleotides resembles the selection of binding proteins by phage display. A large pool of synthetic oligonucleotides (10^{14}–10^{15} molecules) is used for selection. The pool typically has a random region of 30–80 nucleotides and is flanked by two fixed primer sequences. After selection, the resulting aptamers can be shortened in additional work to find the minimal sequence required for target binding, which is usually 15–60 nt (Figure 1.5).

During each selection round, target-binding oligonucleotides are separated from those that bind the target with lower affinity or not at all. Efficient enrichment of the best binding sequences is essential, and several SELEX strategies have been developed to achieve this, including affinity chromatography [305], capillary electrophoresis [177] and target-coated magnetic beads [268] (for a more complete overview of SELEX strategies, see [269]). Selection strategies depend on the desired affinity and specificity of the aptamer, the properties of the nucleic acid (RNA or DNA) used during selection, and

also the properties of the target (e.g. size, charge or presence of functional groups). In addition, aptamers with different features can be selected: e.g. aptamers that are very specific for a given target and do not bind any similar molecules [118], aptamers that self-report target binding [120, 206], or aptamers that catalyze a chemical reaction (ribozymes) [305]. However, incorporation of such features requires a well-designed selection strategy.

Sequences are mixed with target and those bound to the target are recovered and subsequently PCR-amplified. For selection of DNA aptamers, the two strands of the PCR products are separated and one strand is recovered. For selection of RNA aptamers, transcription of the DNA is required before selection, and reverse transcription of the bound fraction is carried out before PCR amplification. In both cases, the resulting enriched pool is the starting material for the next round of selection.

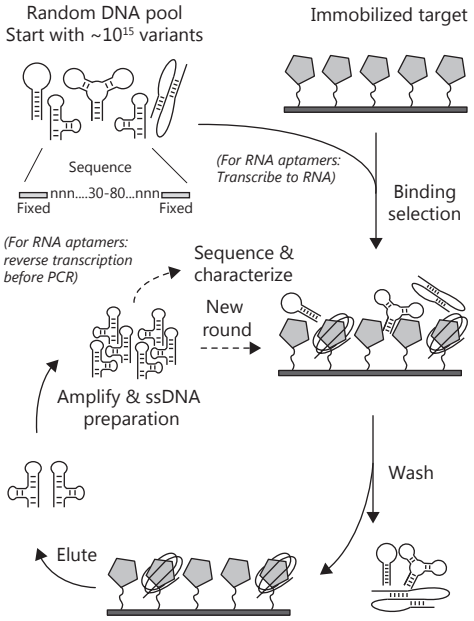


Figure 1.5: Selection of DNA aptamers. A library of random DNA fragments, with a random and two fixed regions, is exposed to an immobilized target molecule. Non-binding DNA molecules are washed away. Bound DNA molecules are eluted, PCR-amplified and made single-stranded. For selection of RNA aptamers, the DNA pool is transcribed to RNA before target binding, and the bound RNA sequences are reverse-transcribed before PCR. The selection and amplification process can be repeated as necessary using more stringent washing conditions in order to obtain higher-affinity molecules. Finally, the DNA molecules can be sequenced and characterized further.

After eight to fifteen selection rounds, remaining sequences are cloned and sequenced. In some rare cases, the pool of potential aptamers is dominated by just one sequence [305]; in most cases, however, a number of different sequences that all fold in similar ways are present [14, 127]. After sequencing, further characterization of several individual sequences is needed to determine the binding kinetics and specificity of binding, to select the aptamer that is most suitable for a predefined application.

The versatility of aptamers

Like antibodies and engineered binding proteins, aptamers can specifically bind targets of interest. Unique features of aptamers, however, are their small size (15–60 nt equals 5–20 kDa; antibodies are 150–160 kDa) and the relative ease of chemical modification (e.g. biotinylation or addition of fluorescent labels). These properties make them useful for a wide variety of applications. Aptamers can, for

instance, substitute for antibodies in the frequently used ELISA assays [55], and they have potential for generating different types of biosensors (see below).

One feature of aptamers is of particular interest for the development of new types of biosensors: upon target binding, the structure of an aptamer changes to some extent. In contrast with antibodies, this structural change can be used to design aptamer sensors that are similar to molecular beacons [288], using a fluorophore-quencher pair at the ends of the molecule, or by incorporation of fluorescent base pair analogues in the sequence [120, 166]. As a result of the structural changes induced by target binding, the fluorophore moves away from the quencher. Consequently, fluorescence is no longer absorbed by the quencher and this can be measured. Following the same principle, electrochemical detection of target binding is possible [3]. Measuring target binding as a result of structural changes is impossible for antibodies, because any structural changes are very small and it is difficult to incorporate a fluorophore-quencher pair within an antibody [155].

Unlike antibodies, aptamers are easily regenerated, at high temperatures or with high salt concentrations, and because of their small size, they can be efficiently immobilized on surfaces with high densities. This allows the design of small and reusable biosensors. Regeneration is also of interest for affinity-purification processes. Aptamers conjugated to resins can, for instance, be used to purify proteins, without a tag, from crude cell extracts [117], or they can be used for the separation of chiral compounds, provided that the aptamer is specific enough for one of the enantiomers [248].

In 2004, Pegaptanib (Macugen) [82] became the first, and currently the only, aptamer-based therapeutic to be approved by the FDA. This RNA aptamer has an antagonistic function, by binding all isoforms, except the smallest ($VEGF_{121}$), of human VEGF, a major regulator of aberrant and excessive blood vessel growth and permeability in the eye. The aptamer does not provoke any undesired immune response because of its nucleic acid origin [196]. However, 2 years later, an antibody-based treatment, Lucentis (antibody fragment), was approved and took over the market, because it can bind all VEGF isoforms, including the smallest. Despite this setback for the aptamer field, Pegaptanib clearly demonstrated the potential of aptamers as therapeutics, and currently several aptamer-based therapeutics are in clinical trials [131]. In addition to therapeutics, aptamers can also be used for targeted drug delivery by coupling them to liposomes [303] or micelles [311] that are loaded with drugs. A recent exciting development in targeted drug delivery is the usage of aptamer-siRNA (small interfering RNA) chimaeras [45, 172, 321]. The aptamer domain of the chimaera specifically binds to a receptor at the surface of a target cell, after which the RNA molecule is imported by endocytosis, and subsequently released from the endosome to enter the RNAi (RNA interference) pathway, via which the siRNA domain ensures silencing of selected target genes.

In the field of biomarker discovery, a new aptamer-based strategy has been described recently [77, 132]. A protein sample is incubated with a mixture of aptamers that contain biotin and a photocleavable group. Aptamer-protein complexes are then captured on streptavidin beads. Once bound, proteins in the complexes are biotinylated and, after washing, the complexes are cleaved from the streptavidin bead. Biotinylated proteins in the complexes are again captured on streptavidin beads; after washing, bound aptamers are eluted and used for hybridization on a microarray. This is a multiplex proteomic approach allowing for high-throughput comparison of proteome profiles, whereas multiplexing in antibody-based methods is very difficult, as described above. This method can potentially be used to select new biomarkers that are specifically related to certain diseases, and to identify potential targets for treatment.

Drawbacks of aptamers

Numerous aptamer sequences for a wide variety of targets have been published, and aptamers have found their way into several applications [131, 163]. SELEX, however, is currently rather labor-intensive, because optimal selection conditions (e.g. selection buffer, number of selection rounds) are target- and application-dependent and need to be optimized for each target. Efforts have been made to automate the SELEX procedure [41], but because optimal selection conditions can vary per target, it is still not straightforward to select a functional aptamer. It is, however, possible to develop aptamers that function under non-physiological conditions. For example, an aptamer against ochratoxin-A that was selected in aqueous buffer, also performed well in the presence of 20% methanol [42].

Potential commercial developments using aptamers are slowed down because only a few companies possess the IP pertaining to aptamers and SELEX technologies. In 1999, Gilead Science Inc. acquired many patents related to general aptamer-selection strategies and of many individual aptamers. Archemix is now exclusively licensed to use the IP for therapeutic applications, and SomaLogic for diagnostic applications. This centralized IP position does not leave room for many competitors, but it should be noted that Archemix has issued commercial therapeutic licences to several companies. In the coming years, however, some patents will expire, allowing more companies to develop aptamer-based applications, possibly resulting in a boost of commercial applications in the near future [179, 284].

In contrast with antibodies, aptamers are small and very stable; the latter feature is especially true for those made of DNA. Nevertheless, aptamers made of the naturally occurring four bases are not fully resistant to degradation; especially in blood serum, unmodified aptamers are rapidly degraded or filtered out of the bloodstream. Modifications in the DNA or RNA backbone or modified nucleobases are often used to make aptamers less susceptible to degradation by nucleases [168].

Modifications can be incorporated either chemically after selection or enzymatically during selection. The most widely used modifications are those that replace the 2′-OH of the ribose sugar, for instance by -NH_2,-F and -OCH_3; these modified nucleotides are all compatible with the enzymatic steps in the SELEX procedure [130, 168]. Frequently used post-selection modifications include: (i) addition of 3′-end caps that reverse the polarity of the oligonucleotide (avoiding 3′ exonuclease activity), and (ii) addition of PEG [poly(ethylene glycol)] molecules at the 5′-end of the aptamer, to fine-tune their *in vivo* half-lives [131]. A completely different approach to prevent degradation makes use of non-naturally occurring L-RNA aptamers, or Spiegelmers [137]; owing to their non-natural origin, they are not recognized by RNases and hence they are very resistant to degradation. In addition, modified nucleotides are used to increase the chemical diversity of the random pool, which is otherwise limited to four bases. This has recently allowed for the selection of aptamers for targets against which no aptamers could be selected with an unmodified pool [290].

Molecular imprinted polymers

A MIP is a polymer that is made in the presence of a template molecule, which is extracted afterwards, thus leaving complementary cavities behind. The procedure to produce these cavities by molecular imprinting was originally published in 1949 [52]. Unlike the above-described affinity tools, MIPs are entirely synthetic scaffolds that have no biological origin. The imprinted polymer material is able to recognize a single target molecule, varying from small organic molecules to proteins, or a group of structurally related molecules based on their shape, size and chemical functionality.

The process of molecular imprinting

Molecular imprinting takes place in several steps (Figure 1.6). It starts with selection of a template molecule that resembles the final target, and of functional monomers based on their ability to interact with functional groups of the template molecule. Functional monomers bind to the template molecule during a pre-polymerization step and are, subsequently, cross-linked to form the imprinted polymer. Finally, the template molecule is removed from the matrix, leaving behind a cavity complementary in size and shape to the template. The obtained cavity can work as a selective

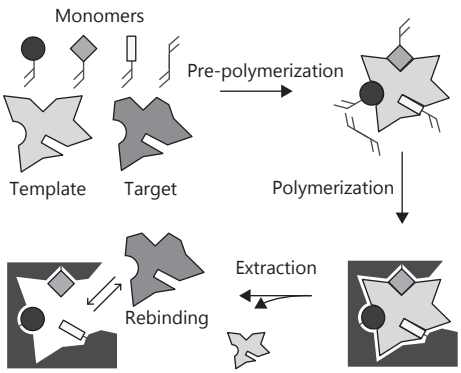


Figure 1.6: Molecular imprinting. Suitable monomers, and a template molecule that resembles the target molecule, are selected and mixed together for pre-polymerization. After polymerization, the template molecule is extracted, and the resulting cavity can be used for target binding.

binding site for a specific molecule.

Two major imprinting approaches can be distinguished on the basis of the interaction between the functional monomer and the template molecule. During imprinting, the template and monomers are linked either by covalent or by non-covalent interactions, such as hydrogen-bonding and Van der Waals interactions. Covalent imprinting generally produces more homogeneous binding sites (analogous to mAbs); however, template removal is often difficult. Non-covalent imprinting generally produces more heterogeneous binding sites, with a rather wide affinity range for the target (analogous to pAbs). Non-covalent imprinting is used most often nowadays, mainly because the process is simple and a wide variety of monomers are commercially available. Design of better monomers and optimized polymerization conditions for the non-covalent imprinting approach has resulted in the production of more homogeneous binding sites [169]. Selectivity and binding kinetics of MIPs need to be determined after imprinting, using the MIP as stationary phase in HPLC columns, or by performing batch rebinding studies [265].

MIPs can be made for small organic molecules as well as for proteins. Conditions for protein imprinting, however, are more restricted. Polymerization can only take place in the aqueous phase, because of limited protein stability in organic solvents. Protein imprinting has been studied intensively, and several imprinting strategies have been developed [69, 256]. The different protein-imprinting strategies can be distinguished into three categories on the basis of the parts of the protein that are imprinted [69, 280]: (i) bulk imprinting, in which the whole protein is imprinted; (ii) surface imprinting, in which only a part of the surface of the protein is imprinted; and (iii) epitope imprinting, where only a small, but typical, part of the protein is used for imprinting. MIPs produced by epitope imprinting can recognize the whole protein by binding a specific part of it, analogous to the way a mAb recognizes an antigen.

MIP composition and shape are important

Molecular imprinting looks deceptively easy, but, in fact, many aspects can influence the final properties of the MIP. Choice of template and functional monomers [125] obviously influence the end product, but the ratio of these molecules is also important. Porosity of MIPs depends on the solvent and temperature used for polymerization. Porosity is important, because pores that are too small limit flow through or make cavities inaccessible, whereas pores that are too large limit the overall binding capacity of the MIP. It is difficult to screen a large number of different polymerization conditions, by conventional techniques, to find the optimal conditions. Several routine conditions for molecular imprinting have therefore been developed over the years, although these might not be the best conditions to make a MIP for any specific target [5]. A semi-automated combinatorial imprinting strategy for preparation of MIPs under different conditions, and subsequent evaluation of these MIPS, has been developed and may

shorten the time required to determine optimal polymerization conditions [279].

Conventional methods for MIP production result in large polymer blocks; subsequent grinding and sieving yield similarly sized, but irregularly shaped, particles. Irregular particles are not a problem for certain applications [e.g. solid-phase extractions (SPEs)], but some other applications require size- and shape-defined particles (e.g. *in vivo* applications). Size and shape-defined MIPs can nowadays be produced in several different formats such as bead, membrane, monolayer, nanowire or dendrimer [314].

MIPs work well in non-aqueous environments

MIPs are very stable because of their synthetic nature; they can resist high temperatures and a wide pH range, and can be stored for several years without losing affinity [70]. Unlike the affinity tools described before, MIPs work especially well in organic solvents instead of aqueous solutions. Production of MIPs is cheap and fast, once appropriate functional monomers and cross-linkers have been chosen. Numerous imprinted polymers, capable of binding toxins (e.g. Bisphenol A), pesticides (e.g. catechol), antibiotics (e.g. penicillin G) are now commercially available (see companies in Table S1.1), and some companies even offer to develop custom-made MIPs.

Imprinted polymers are most frequently used in SPE, a separation process that is used to concentrate target molecules or to remove contaminants during downstream processing. SPE is often used for sample preparation before analysis [287], for instance for the separation of enantiomers, or concentration of analytes from waste water or food samples. To ensure effectiveness of SPE, however, conditions should closely resemble those originally used to prepare the MIP. Even when using these conditions, extensive optimization of the SPE procedure is often required [287]. Most MIPs are prepared in organic solvents, hence it is difficult to use aqueous solutions directly [33]; conventional SPE can be used in this case to exchange aqueous solutions for those that more closely resemble polymerization conditions.

Application of MIPs in a molecular-imprinted sorbent assay, analogous to ELISA, for detection of theophylline and diazepam was reported for the first time in 1993 [291]. Binding affinity of target molecules to MIPs was deduced from a competition assay with non-labelled and radiolabelled target. Results and reproducibility were good compared with existing assays, but the procedure is more time-consuming and requires the synthesis of radiolabelled target compounds. Although this assay does not proceed well in aqueous solutions, it does work very well in organic solvents. In addition to these molecular-imprinted sorbent assays, MIPs are currently also used in chemical sensors; with different modes of detection, e.g. electrochemical, optical and mass-sensitive [98].

The application of MIPs in the drug discovery process was described previously [223]. Molecular imprints of receptor-binding molecules are made and, subsequently,

used for screening compound libraries to identify molecules that have agonistic or antagonistic properties [30]. Imprinted cavities can alternatively be used as template for synthesis of compounds that fit within the cavity, mimicking the template molecule [218, 317].

Few reports about *in vivo* applications of MIPs have been published, but efforts have been made to use MIPs for adsorption of cholesterol from body fluids [112]. MIPs have only recently been successfully used for absorption of melittin, which induces cell lysis at high concentrations in the bloodstream of living mice [104]. However, as the majority of MIP applications are restricted to non-aqueous environments, *in vitro* applications will be the main area of use for MIP technology.

The most attractive affinity tool

In the previous sections, the production, selection and relevant characteristics have been described for antibodies, engineered binding proteins, aptamers and MIPs. However, an important question remains unanswered: which tool has the most attractive features? This question is quite difficult to answer, because there is not a single affinity tool that fits all of the requirements for all of the applications. Each affinity tool has its own specific characteristics (see Table 1.1) and, depending on the intended application, one tool has more potential than another. In this section, we discuss the applicability of each affinity tool and relate this to various applications.

Table 1.1: Affinity tool characteristics.

Characteristics	Antibody	Binding protein	Aptamers (ssDNA and RNA)	MIPs
Size (kDa)	~150–160	<30	5–20 (15–60 nucleotides)	n.a.
Selection	In vivo	In vitro	In vitro	In vitro
Production	Animal or recombinant	Recombinant	Synthetic	Synthetic (large scale low costs)
Post-selection modifications	Possible, but heterogeneous products	Possible, can be designed for homogeneous products	Wide variety of options (sugar, base, or phosphate; 5', 3', or internal)	Possible, but should be taken into account during imprinting
Stability	Several weeks at 4°C	Variable	DNA: Years at room temperature. RNA: several months at -80° C	Years at room temperature
Binding site	Monoclonal: homogeneous Polyclonal: heterogeneous	Homogeneous	Homogeneous	Depending on imprinting strategy, heterogeneous or homogeneous
Target molecules	Mainly immunogenic macromolecules	Macromolecules and low molecular weight molecules	Low molecular weight molecules, macromolecules and cells	Mainly low molecular weight molecules (<1000)
In vivo half-lives	Days to weeks	n.a.	Untreated: seconds - minutes. Treated: days	n.a.
Application conditions	Physiological	Physiological and non-physiological	Physiological, non-physiological and organic solvents (to some extent)	Mainly organic solvents, aqueous solutions to some extent

n.a., not applicable

Affinity chromatography

The purification of proteins or small molecules by affinity chromatography is widely applied in all areas of life sciences and biotechnology. The process relies on purifying the target protein or molecule of interest from a complex mixture of molecules, on the basis of the target's specific and high-affinity binding to an immobilized ligand, resulting, in the ideal case, in a pure target in a single step.

Antibodies are currently used for antigen purification, immunoprecipitation and pull-down assays; however, they are rarely used for downstream processing of proteins or other molecules on an industrial scale. In general, large quantities of the affinity ligand are required to construct such columns. In addition, antibodies have a limited stability, a large size and high production costs, and are therefore not the first choice for this application. Recently, however, a new development in immunoaffinity chromatography was reported, where an anti-caffeine camelid antibody was used for the separation of caffeine and related methylxanthines [62].

Aptamers are very promising in the field of affinity chromatography [39]. Human L-selectin was purified from CHO cell-conditioned medium using a DNA-aptamer immobilized as an affinity chromatography matrix [241]. More recently, Taq DNA polymerase was purified using a DNA-aptamer immobilized on magnetic beads [207]. However, despite their promise as tools in affinity chromatography, there have been relatively few examples of the use of aptamers [316]. This may, at least in part, be due to their variable stability and their sensitivity to nuclease activity. However, it can be expected that stabilized non-natural nucleotides will circumvent the instability problem of natural oligonucleotides, and that aptamers will find a wider application as ligands for protein or small-molecule purification in the near future.

MIPs have specifically received interest for the extraction of small molecules from complex food and other biological samples. There are various examples of successful MIPs for relatively low-molecular-mass molecules, such as antibiotics and toxins [323]. MIPs are very well suited for SPE of small molecules; for example, a range of highly selective SPE phases based on MIPs have recently become commercially available (Biotage AB). However, MIPs, are at this moment, not suitable for protein purification purposes because of their typically harsh operational conditions.

Engineered binding proteins are potentially the most attractive alternative to antibodies in affinity chromatography. They are more likely to withstand the harsh column washing and regeneration conditions and are cheaper to produce due to their smaller size and simple scaffolds. At this moment, the most frequent use of affinity chromatography using an engineered binding protein is the purification of antibodies using recombinant Protein A or Protein G [92]. Recently, a number of Affibody proteins coupled to affinity resins were demonstrated to specifically capture their target molecule from complex mixtures [83]. In addition, Affibody proteins have been included in commercially available depletion columns from Agilent Technologies in order to

remove interfering high abundant proteins from human samples (e.g. plasma, serum). These results strongly suggest great potential for the development of engineered binding protein resins for the purification of specific proteins and other molecules.

Diagnostics

Antibodies are at this moment the best and most frequently applied tool for routine diagnostic tasks. ELISA and Western blot analyses are used for analyses in a range of medical applications, but also for analyzing food and environmental samples. In addition, despite their relatively high costs and limited stability, antibodies are increasingly used as point-of-care diagnostics, such as pregnancy and HIV tests [20, 293]. Commercial availability is an important contributor to the widespread use of antibodies; it compensates for their relatively high price and limited stability to some extent. Development of cheaper or more stable alternatives is to be expected in the future. However, actual implementation of an alternative affinity tool for an existing application has often proven to be difficult. Once a procedure has been established and validated, the advantages of using an alternative affinity tool must be worth the effort of repeating all of the laborious and expensive testing to validate the procedure. Alternatives for antibodies, e.g. MIPs [291] and aptamers [170], have indeed been developed for ELISA assays; however, because of the above-mentioned reasons, they are not yet widely applied. Antibodies remain the most attractive tool for routine diagnostic tests. However, in case an assay should be performed at non-physiological conditions, or in organic solvents, aptamers, MIPs or engineered binding proteins are certainly worthy of consideration as a binding tool.

Therapeutics

When an affinity tool has to be developed from scratch, the characteristics of each tool and prerequisites for the intended application should be evaluated carefully in order to select the most appropriate tool. Requirements for human *in vivo* applications, for example, are the most demanding, in terms of legislation and complexity of the conditions in which the tool should operate. This is probably why antibodies have received most attention, because they can be selected and produced in an environment that closely resembles human *in vivo* conditions. Over the years, antibodies have come a long way from their initial *in vivo* selection and production, and have become very successful especially as therapeutics; many problems concerning immunogenicity have recently been solved [9, 296]. However, as antibodies also have their limitations, alternative tools have certainly gained interest; one aptamer has been FDA-approved, and several aptamers and several engineered binding proteins are currently in clinical trials [9, 130]. In this respect, MIPs receive less attention, despite the recent development of using them for detection and binding of melittin *in vivo* [104]. Before MIPs can enter clinical trials, several key issues need to be addressed, such as compatibility of MIPs

with aqueous solutions.

Although the potential for *in vivo* application of aptamers and engineered binding proteins is clearly demonstrated, it will take many years before they can compete with the success of antibodies. For antibodies, however, it also took years before they became successful, perhaps best illustrated by the 8 years it took before a second antibody therapeutic was FDA-approved, in 1994, after approval of the first antibody therapeutic, in 1986.

Outlook

At present, antibodies are the most popular and best established affinity tools, especially in diagnostics. It appears very unlikely that alternative affinity tools will play a significant role in the field of diagnostics soon, simply because of the wealth of antibody-based assays that are readily available. Still, there may be exceptions. The alternative affinity tools are, at this moment, more attractive for applications in which antibodies are not well suited, for reasons that relate to size, physical and chemical stability, production costs and requirement for high-throughput analysis. For instance, affinity chromatography and SPE generally require cheap and stable affinity tools. Antibodies are too expensive and too unstable for most of these applications; hence application of engineered binding proteins, aptamers or MIPs might be more attractive in these cases. Moreover, aptamers and engineered binding proteins are both very promising for *in vivo* imaging, because they are much smaller than full-length antibodies, and hence can better penetrate tissues, resulting in increased signal to noise ratios. It is important to note, however, that, although diverse applications may require different affinity tools (Table 1.2), the most attractive tool for a certain application may not necessarily be the one with the best biochemical and physical features. Obviously, economic properties, such as the commercial availability and ease of selecting improved variants, are also very important considerations.

Table 1.2: Suitability of affinity tools for various applications.

Application	Antibodies	Engineered binding proteins	Aptamers	MIPs
Therapeutics/ treatment	++	+	++	
Targeted drug delivery	++	+	+	
Molecular imaging	++	+	+	
Drug discovery				+
Diagnostics	++	+	++	+
Affinity purification	+	++	+	++
Biosensors		+	+	++

+, reported in literature; ++, commercially available

As research on aptamers, engineered binding proteins and MIPs continues, it is anticipated that, in the cases of the aforementioned niche applications in which antibodies perform poorly, they will either progressively replace antibodies or enable completely novel applications. It is tempting to speculate that the types of affinity ligands discussed will eventually become readily available for routine diagnostic and industrial applications, as a useful complement of the currently available antibodies.

Funding

This project was financially supported by grants from NWO (Netherlands Organization for Scientific Research)/SRON (Netherlands Institute for Space Research) [User Support Space Research Programme contract ALW-GO-PL/08-08], NIVR (Netherlands Agency for Aerospace Programmes) PEP (projects APTALIFE and APTALIFE DEMO), and The Dutch Separation Technology Institute (DSTI).

Supplementary information

Table S1.1: Selected companies developing affinity tools based on novel scaffolds. For a more extensive list of therapeutic antibodies, see [195, 224].

Affinity tool	Company	Acronym	Scaffold	www address
Engineered affinity proteins	Adnexus (acquired by Bristol-Myers Squibb)	AdNectins	Human tenth FN3	http://www.adnexustx.com
	Affibody	Affibodies	Protein-binding domain of Staphylococcus aureus protein A	http://www.affibody.com
	Amgen (formerly Avidia)	Avimer	LDL (low-density lipoprotein) cell-surface receptor domain A	http://wwwext.amgen.com
	Arana (formerly EvoGenix Therapeutics)	Evibodies	Human cytotoxic-associated antigen (CTLA-4)	http://www.arana.com
	Pieris AG	Anticalins	Lipocalins	http://www.pieris-ag.com
	BioTech Studio	Telobodies	Human PDZ domains	http://www.biotechstudio.com
	Bionexis	AFIM	Phospholipid interaction domain of annexin	http://www.bionexis.com
	Borean Pharma	Tetranectins	Human C-type lectin domain (CTLD)	http://www.boreanpharma.com
	Covagen	Fynomer	Fyn SH3 (Src homology 3) domain	http://covagen.ch
	Dyax		Kunitz domain of trypsin inhibitors	http://www.dyax.com
	Immunocore (formerly Avidex)	mTCR	Monoclonal T-cell receptors	http://www.immunocore.com
	Molecular Partners	DARPins	Ankyrin repeat protein	http://www.molecularpartners.com
	NascaCell (formerly Selecore)	Microbodies	Cysteine knots or knottins	
	Pacific Northwest National Laboratory, Department of Energy	Top7		http://www.pnl.gov
	Scil Proteins	Affilin	Human ubiquitin/human γ -crystallin	http://www.scilproteins.com
	Emergent Biosolutions (formerly Trubion Pharmaceuticals)	SMIPs	Binding domain from annexin	http://www.truemergent.com
Aptamers	AdAlta		Peptide aptamers	http://www.adalta.com.au
	Aptagen	Aptabody	Nucleic acid aptamers	http://www.aptagen.com
	AptaRes	MonoLex	Nucleic acid aptamers	http://www.aptares.net
	Archemix		Nucleic acid aptamers	http://www.archemix.com
	Imaxio		Peptide aptamers	http://www.imaxio.com
	NascaCell (formerly Selecore)		Nucleic acid aptamers	http://www.nascacell.de
	NOXXON Pharma AG		Nucleic acid spiegelmers	http://www.noxxon.com
	SomaLogic	SOMAmers	Nucleic acid aptamers	http://www.somalogic.com
MIPs	Biotage AB (formerly MIP Technologies)	AFFINILUTE	Molecular imprinted polymers	http://www.biotage.com
	PolyIntell	AFFINIMIP	Molecular imprinted polymers	http://www.polyintell.com
	Detection Inc.		Molecular imprinted polymers	http://www.raptordetection.com
	Semorex		Molecular imprinted polymers	http://www.semorex.com

Chapter 2

Initiating Surface Plasmon Resonance experiments and processing kinetic data

Vincent J.B. Ruigrok, John van der Oost, Hauke Smidt

Data presented in this paper were obtained within the framework of the 2011 Global Label-free Interaction Benchmark study, initiated by Rebecca L. Rich & David G. Myszka, Biosensor Tools, LLC; Salt Lake City, USA.

Abstract

Virtually all interactions of biomolecules can be studied using a Surface Plasmon Resonance (SPR)-based sensor. Such a sensor can monitor binding interactions in real-time, which makes it possible to study the actual kinetics of complex formation and dissociation. Obtaining high quality kinetic data, however, is usually not straightforward and requires optimization of experimental conditions. Despite the efforts that have been made to educate users of SPR-based biosensors on how experiments should be performed correctly, a lot of the published SPR data is still of poor quality at best. In this chapter, experimental considerations involved in initiating and properly performing SPR experiments as well as data analysis and presentation are discussed.

Introduction

Interactions of biomolecules are widespread and very diverse in nature. They can, however, virtually all be studied using a Surface Plasmon Resonance (SPR)-based sensor. With such a sensor, binding interactions can be followed in real-time, which makes it possible to study the actual kinetics of complex formation and dissociation. In addition, SPR does not require any labelling of the binding partners; this is advantageous, because labelling can adversely affect the interaction. Besides SPR there are other label-free techniques that can be used to study binding interactions, such as Isothermal Titration Calorimetry (ITC) [63, 73], or Microscale Thermophoresis [119, 302]. These techniques, however, can only be used to perform equilibrium assays rather than measuring kinetics. Knowledge about underlying kinetic parameters is certainly relevant, not only for better understanding of the molecular basis of complex formation, but also because different kinetics can result in a similar affinity (Figure 2.1). Such largely different interactions, from a kinetic point of view, would be considered equal when measured using an equilibrium assay. Alternative methods that are capable of measuring kinetics, on the other hand, are not label-free, such as Kinetic Exclusion Assays (KinExa) [44], or stopped-flow assays [49].

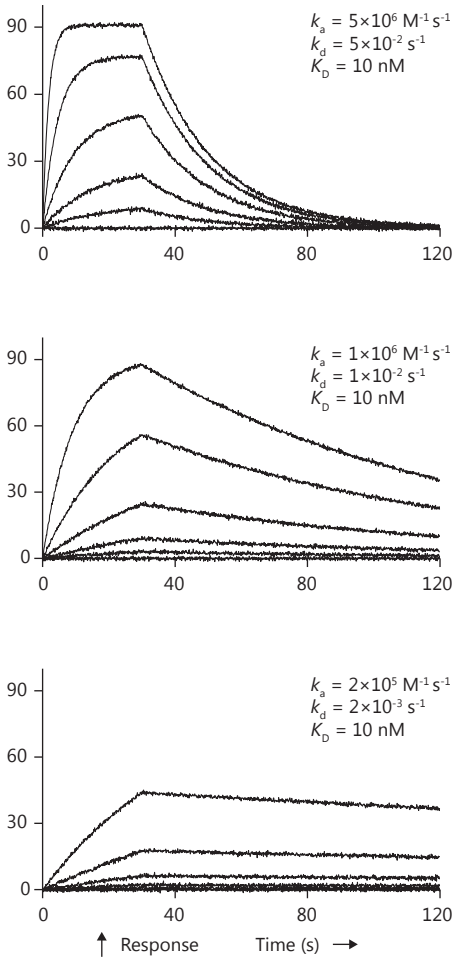


Figure 2.1: Simulated kinetic data. All three data sets have identical affinities (K_D =10 nM), but widely different kinetics (k_a, k_d). A maximum binding capacity of 100 Response Units (RU) and analyte concentrations of 0, 1.2, 3.7, 11.1, 33.3 and 100 nM were assumed for each simulation. The program CLAMP was used for these simulations [193].

Despite the diverse nature of interactions, the majority of the interactions studied using an SPR-based biosensor can be described by a simple bimolecular interaction model (equation 2.1), if the experiment and data analysis are done properly [232], even though other, more complex models, have been described and used as well [122, 185].

Equation 2.1: $$A + B \underset{k_d}{\overset{k_a}{\rightleftarrows}} AB$$

In which the association rate constant k_a ($M^{-1}\,s^{-1}$) is a measure for the rate of complex formation, and the dissociation rate constant k_d (s^{-1}) is a measure for the rate at which the complex falls apart. The equilibrium dissociation constant, K_D (M), of this reaction is a measure for the overall strength, or affinity, of the interaction and can be obtained in different ways (Equation 2.2).

Equation 2.2: $$K_D = \frac{k_d}{k_a} = \frac{[A] \times [B]}{[AB]}$$

In which k_a and k_d are the rate constants described above, [A] and [B] are the concentrations of the two individual binding partners, and [AB] the concentration of the complex at equilibrium conditions.

In depth knowledge of the physical phenomenon of SPR [99, 100, 109] is not required in order to operate the biosensor. It is therefore not without reason that the interest in SPR-biosensors has been growing ever since the introduction of the first commercial SPR sensor in 1990. Numerous SPR-based systems are available nowadays, but Biacore (GE Healthcare) is still by far the main supplier [239]. The increased use of SPR-based biosensors is not merely positive, as a lot of the published SPR data is of poor quality at best [233]. This seems surprising, because a large effort is made to educate users of SPR-based biosensors on how experiments should be performed correctly and to stimulate them to actually do so. A wealth of information on various aspects of SPR-based biosensing is available in the (annual) surveys of the biosensor literature [229, 230, 233, 235], and in various other SPR related publications [189, 231, 232, 234, 239]. The primary aim of this chapter is therefore not to provide a complete overview, but rather to serve as a starting point for those who wish to train themselves in the art of SPR-biosensing; by providing an easy accessible overview on the basics of SPR and elaborating on the experimental considerations involved in initiating and properly performing SPR experiments. The focus will be on experiments that aim to obtain kinetic and affinity parameters of binding interactions. Data analysis and presentation are important factors that are often neglected in literature [233], and will therefore also be discussed here.

General principle of an SPR-biosensor

Polarized light that propagates in a medium of high refractive index (e.g. glass), can be fully reflected if it encounters a medium of lower refractive index (e.g. sample solution) at a specific angle (Figure 2.2); this phenomenon is called total internal reflection [239]. Under conditions of total internal reflection an electromagnetic field 'leaks' into the medium with low refractive index, resulting in an evanescent

wave. The amplitude of the evanescent wave decreases exponentially with the distance from the surface. If the interface between the two media is coated with a thin layer of metal (usually gold), the reflected light will show a dip in intensity at a certain angle of reflection: θ (for more details see [99, 109, 239]). The angle at which the dip in intensity occurs is very sensitive for the refractive index of the solution within the penetration depth of the evanescent wave. Analytes binding to ligand molecules within the evanescent field cause a change in refractive index and hence in the angle of the intensity dip (θ). In Biacore systems the angle at which the intensity dip occurs is monitored as a measure of binding.

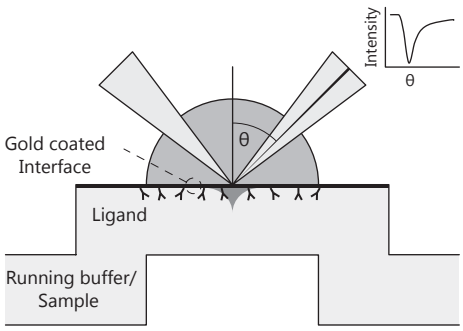


Figure 2.2: General principle of SPR. In this figure the dip in reflected light intensity is indicated with θ, see text for details.

Although SPR is label-free, it requires that one of the binding partners is immobilized; this requirement is an often heard criticism, because as a result the measured values might not reflect the actual binding behavior in solution. However, over the years several studies have been published comparing SPR with various solution based methods. Their results clearly show that, when each technique is used properly, they produce comparable results [2, 49, 54]. In part this is because molecules are often not directly attached to the surface, but to a matrix (dextran) that is attached to the surface.

Reagents

For all types of interactions studied using SPR, high quality reagents are essential. Both analyte and ligand should (proven to) be highly pure and in a low polydispersity state. Data quality is linked to reagent quality, therefore, and despite all the experimental conditions that can be adjusted, low quality reagents will never yield high quality data.

Chips

Chips with a variety of surfaces, with different characteristics, are commercially available nowadays for Biacore systems; two suppliers are Biacore (GE Healthcare, Sweden) and XanTec bioanalytics GmbH (Germany). Most chips, however, have a dextran hydrogel deposited on a gold layer. This dextran layer provides a hydrophilic environment and an increased surface capacity for ligand binding. Biacore's CM5 chip (XanTec equivalent: CMD500L) is a multipurpose chip and a useful choice to start experiments, although certain experiments may require more specialized surfaces [1].

Coupling vs capture

As described before, SPR analysis requires one of the interacting molecules to be immobilized on the chip surface, either via covalent coupling (Figure 2.3A) or via non-covalent capture approaches, using a covalently-bound capture molecule (Figure 2.3B). Covalent attachment usually provides a more stable surface compared to capturing. Nevertheless, despite that capturing is often less stable, it is done at milder conditions which decreases the risk of ligand inactivation.

Covalent attachment

Covalent attachment of ligands can be done, amongst others by using primary amines, carboxylic acids or sulfur groups present in the ligand [1]. The most straightforward, and most frequently used approach, is ligand coupling via a primary amine in the ligand [using N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC) and N-Hydroxysuccinimide (NHS)] [236]. The negatively charged carboxyl groups in the dextran layer are essential, as they allow electrostatic attraction of positively charged protein ligands to the surface (pre-concentration). Pre-concentration allows for efficient coupling of protein ligands from dilute solutions (10–50 µg ml^{-1}); usually pre-concentration of a protein ligand should be tested by diluting it in low salt buffers (10 mM sodium acetate) with varying pH, usually 4.0–6.0. The pH that results in the most efficient pre-concentration is not necessarily the best for coupling [22]; especially when a low density surface is required (discussed below), or when the protein is expected to be unstable at low pH. In these cases one might choose a higher pH at which the ligand is not as heavily pre-concentrated at the surface. It is worthwhile to keep in mind that the coupling reaction with EDC and NHS is slow, especially at low pH; for this reason, remaining active groups are often blocked using 1 M ethanolamine at pH 8.0. In addition, pre-concentration is often not as efficient for small molecules. When a small molecule is used as ligand it is recommended to use a higher ligand concentration (µM–mM range, depending on solubility and surface density desired) and perform the coupling reaction at neutral pH.

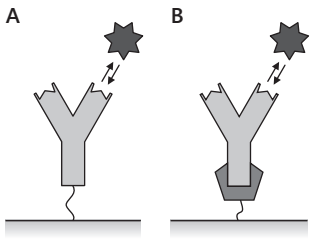


Figure 2.3: Immobilization methods. A) Covalent attachment of the Y-shaped ligand molecule to the chip surface. B) The ligand molecule is captured by another molecule (e.g. protein A/G, antibody, streptavidin) that is covalently attached to the chip surface. The analyte is depicted as star-shaped molecule.

Capture

Ligand capture can make use of the various tags that can be added to a protein; a protein with a His tag could be captured on a Ni/NTA surface, or on a surface coated with anti-His antibodies. Antibodies for a variety of purification tags are available;

once coupled to a chip surface these antibodies could even be used to capture ligands from crude extracts. In addition, capture using antibodies produces a homogeneous surface, which could reduce complexity in the data. Alternatively, a ligand could be biotinylated for subsequent capture on a streptavidin surface. The strength of the biotin-streptavidin interaction is often exploited in capture approaches, despite that ligand biotinylation is a random process and therefore produces a heterogeneous surface. Over-biotinylation of the ligand should be prevented, as it can significantly decrease the surface capacity [214]. Addition of an Avi-tag, that allows for site specific biotinylation [10], combines the site specificity of an antibody capture approach and the strength of the biotin-streptavidin interaction.

Deciding which binding partner to immobilize

Binding of heavier molecules to smaller ligands yields bigger –easier to measure– responses, however, coupling the smallest binding partner to the chip might not necessarily be the best approach. Several practical aspects can influence the decision on which binding partner should best be immobilized.

Non-specific binding

Before deciding which binding partner to immobilize it is recommended to check whether they interact with an untreated surface. This can easily be done by injecting the highest concentration of each binding partner that will be used in subsequent experiments. Small amounts of non-specific binding can be compensated for by using a reference channel (discussed below). Otherwise experimental conditions should be altered to minimize non-specific binding. Several additives can be included in the running buffer to reduce non-specific binding. Examples are the addition of more salt, detergent (usually 0.05% of Surfactant P20 or Tween 20), BSA (anywhere between 0.1 mg ml^{-1} to 20 mg ml^{-1}) and soluble dextran (~1 mg ml^{-1}). Depending on the sample type, some approaches might be more effective than others, therefore it is advisable to assess effectiveness of additives in preventing non-specific binding. In addition, different chip surfaces might be explored. If non-specific binding cannot be reduced to a satisfactory level, the most sticky protein should be immobilized.

Avoiding avidity

Avidity effects are especially a concern when studying the kinetic behavior of bivalent (or multivalent) molecules, such as antibodies. A bivalent molecule in solution could potentially bind two ligands on the chip surface (Figure 2.4A). This will result in a higher apparent affinity, and the resulting data cannot be described with a simple 1:1 model. A possible strategy might be to couple only a small amount of ligand to the surface, preventing that a second ligand is within reach, once the bivalent molecule has bound one ligand molecule (Figure 2.4B). The easiest solution to avoid avidity,

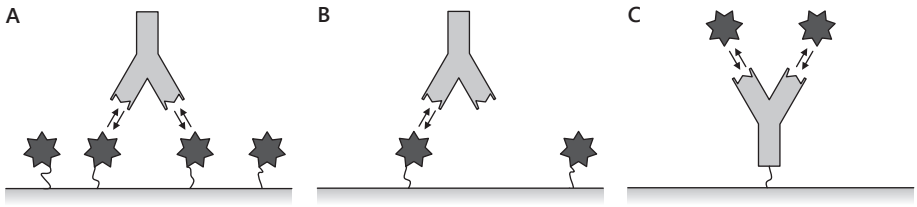


Figure 2.4: Reducing avidity effects. A) Avidity effects can arise when a bivalent molecule is injected over a surface that contains a high density of ligand. B) Avidity effects can be reduced by binding less ligand on the surface. C) The best way to avoid avidity is to attach the bivalent molecule to the surface. Bivalent molecules are depicted as Y-shaped molecules, monovalent molecules are depicted as star-shaped molecules.

however, is to couple the bivalent molecule to the chip surface and inject the other binding partner (Figure 2.4C).

Collecting high quality data

Obtaining high quality data requires more than good reagents and sensible decisions on which molecule should be attached to the chip surface. Experimental conditions, for instance the use of a reference surface, flow rate and regeneration can improve data quality to a large extend as well.

Reference surface

Data quality can be greatly improved by using a reference surface, as it can be used to correct for bulk contribution, low levels of non-specific binding, injection noise, and baseline drift [183, 191]. Using a reference surface is straightforward in instruments that either have multiple flow cells (e.g. Biacore 2000 and 3000) or an array (e.g. Biorad ProteOn XPR36, IBIS MX96), because the analyte passes over the reference and analyte surfaces simultaneously. To ensure a similar environment, a reference surface is best treated in the same way as the ligand surface. The best reference surface contains a modified version of the ligand that is known not to bind the analyte. When no such modification is available, a similar but non-target protein, or an unrelated protein will suffice. The reference channel should at least be activated and blocked in the same way as the ligand surface.

Blank injections

Blank injections are injections of the same buffer that is used as running buffer. Ideally there should be no detectable response, but often there is a minor change in the bulk refractive index. This change, however, should be consistent between the flow cells and subsequent injections. In the beginning, a number of injections usually deviate from subsequent injections (usually the first and second). Therefore, a number of blank injections (usually >4) preceding analyte injections help to equilibrate the

instrument, resulting in a more stable and reproducible signal. Blank injections are also necessary for double referencing of the data.

Assay types and sample injections

Regardless of the assay type used, it is important to use a dilution series that spans a wide concentration range (typically 100 fold). In addition, it is better to assay a small number of concentrations (generally 5) and replicate them, rather than injecting a large number of concentrations only once. Kinetic measurements can be performed in various assay formats on a Biacore 3000: classical kinetics, short-'n-long and titration kinetics (Figure 2.5). In the classical approach a dilution series of analyte is injected over the surface in subsequent cycles; each cycle takes an equal amount of time. It is essential to have curvature in the data to derive accurate k_a and k_d values, however, especially for interactions with a low k_d it can take a long time to get sufficient curvature in the dissociation phase. The short-'n-long assay is therefore used to save assay time; only the dissociation phase of the highest analyte concentration is monitored long enough to get sufficient curvature in the data. In a titration kinetics approach, a dilution series is injected over the surface during one cycle, starting with the most dilute concentration; there is a short dissociation time between injections. The analyte injections are followed by a single long dissociation phase (analogous to the short-'n-long approach). Titration kinetics reduces the need for regeneration of the ligand surface.

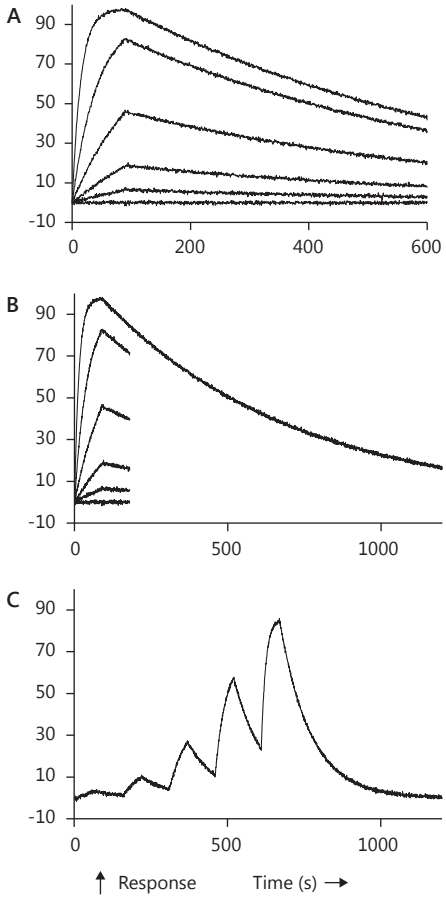


Figure 2.5: Assay formats for kinetic analysis. A) Classical kinetics assay, all analyte injections are monitored for the same amount of time. B) Short-'n-long assay, to save time, only the dissociation phase of the highest analyte concentration is monitored long enough to obtain curvature in the data. C) Titration kinetics, increasing analyte concentrations are serially injected, followed by a single dissociation phase, this assay format reduces the need for regeneration.

Reducing mass transport limitations

Liquid flow in a flow cell is laminar, resulting in an unstirred boundary layer at the chip surface. The thickness of the unstirred layer depends, amongst others, on the flow rate: the higher the flow rate, the smaller the unstirred boundary layer (Figure 2.6). Mass transport limitation occurs in the unstirred boundary layer if transportation of the analyte to the ligand is limited by diffusion. In addition, if a dissociated analyte is not timely removed from the unstirred boundary layer, rebinding events can occur.

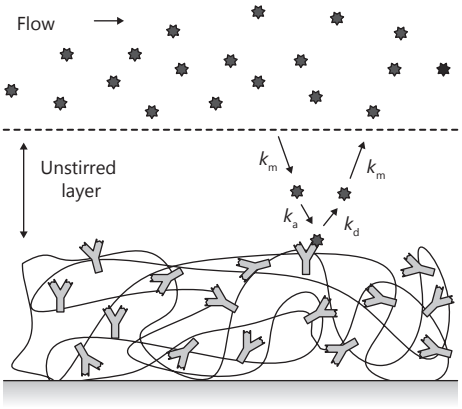


Figure 2.6: Mass transport limitation. This occurs in the unstirred boundary layer if transportation of the analyte to the ligand is limited by diffusion, or if a dissociated analyte is not timely removed from the unstirred boundary layer. The rate at which the analyte is transported in and out of the unstirred boundary layer is described by the diffusion limited rate constant (k_m) [192]. The unstirred boundary layer decreases with higher flow rates.

Mass transport limitations can be reduced by using a high flow rate and a low surface capacity. High flow rates result in a thinner unstirred boundary layer; a low surface capacity reduces potential rebinding events [189]. For kinetic measurements a flow of 30 µl min^{-1} is often used [236], and the highest response should preferably not exceed 100 Response units (RU) [190]. When a binding reaction is limited by mass transport, the association phase shows linearity and has little to none curvature; in addition the observed dissociation is less than the actual dissociation (analogous to avidity effects).

Especially when the association rate (k_a) of a binding reaction is high (above 1×10^6 M^{-1} s^{-1} [239]) it is difficult to fully avoid mass transport limitations. In these cases, a two-compartment model is often used to fit and analyze data. In this model, the simple bimolecular interaction model (equation 2.1) is extended to yield equation 2.3.

Equation 2.3: $$A_0 \underset{k_m}{\overset{k_m}{\rightleftarrows}} A + B \underset{k_d}{\overset{k_a}{\rightleftarrows}} AB$$

The rate at which the analyte is transported in and out of the unstirred boundary layer (the transition between A_0 and A) is described by the diffusion limited rate constant (k_m) (Figure 2.6) [192].

Regeneration

Regeneration is used to prematurely end the dissociation phase by injecting a short pulse of a solution that disrupts the complex. A regeneration step can reduce the assay time considerably, especially if the complex dissociates slowly. The regeneration condition, however, should be such that the analyte is efficiently removed without

adversely affecting the ligand. Although the best regeneration condition should often be determined empirically, knowledge on the interaction and interaction partner might provide a clue on which types of regeneration (e.g. acid, base, SDS or salt) are worthwhile to test. In general, regeneration should be a short pulse (10–20 s) at a high flow rate. In addition, when screening for an appropriate condition it is recommended to start with a mild condition and gradually progress to harsher conditions. Once a suitable condition is found a series of binding and regeneration experiments should be performed, to assess whether or not the ligand is damaged by the regeneration. When running a sequence of injections, it is recommendable to add a buffer injection after the regeneration in order to prevent carryover of the regeneration solution to the subsequent analyte injection.

Analyzing and presenting data

After data collection the raw data needs to be processed such that the binding responses become apparent. Figure 2.7A shows the raw data of a titration kinetics experiment, which was conducted in the framework of the 2011 Global Label-free Interaction Benchmark study (see supplemental text for details). The actual response of the binding event might be difficult to distinguish in the raw data, but becomes apparent after processing.

Processing data

Two higher curves in Figure 2.7A correspond to the ligand coated channel, the bottom curves are from the reference channel. The first step in processing data is to set these responses for all channels to zero on the y-axis. Zeroing is done by subtracting the average response a few seconds before the injection start (Figure 2.7B). The next step, subtraction of the signal of the reference surface, eliminates most of the injection noise and bulk changes (Figure 2.7C). Sharp spikes can arise, because of the different positions of the reference and ligand surface, and should be eliminated from the data. Although the signal of binding (gray) becomes clearly visible the responses still show some bulk changes. These seemingly minor changes can adversely affect the fitting process and should be further corrected by double referencing of the data. Double referencing means that the average signal of several blank injections (black) is subtracted from the data (Figure 2.7D). Subsequently, replicate injections are overlaid and this data can now be analyzed using a titration kinetics model [126] (Figure 2.7E). Curvature is essential to obtain accurate values for the rate constants of association (k_a) and dissociation (k_d). It should be noted, however, that even with lack of curvature, it is still possible to obtain an accurate K_D as it is defined by the ratio $K_D = k_d / k_a$.

Presenting data

Generally it requires considerable effort to obtain high quality data and to extract the

binding constants. However, this is not yet the time to lean back and ponder on the biological implications of the results. Presentation of the data and the fits deserves attention as well! In fact, why would you not show the data, especially considering the efforts it took to obtain it? Not rarely, however, only the binding constants are reported without showing any data [233], which makes you wonder how the experiment was executed.

The data can give a trained eye an indication on how well the experiment was performed and hence how reliable the reported binding constants are. A figure should preferably contain an overlay of all replicate measurements, together with the fit to a 1:1 model (by default). For long measurements it is useful to show the response of one or more blank injections, to assess the stability of the baseline during the measurements. Sharp spikes should be removed as they are not relevant; moreover, spikes force relevant data to a scale that is too small to be useful. In addition, the figure should focus on the binding event, and should preferably show only the baseline just before and just after the injections.

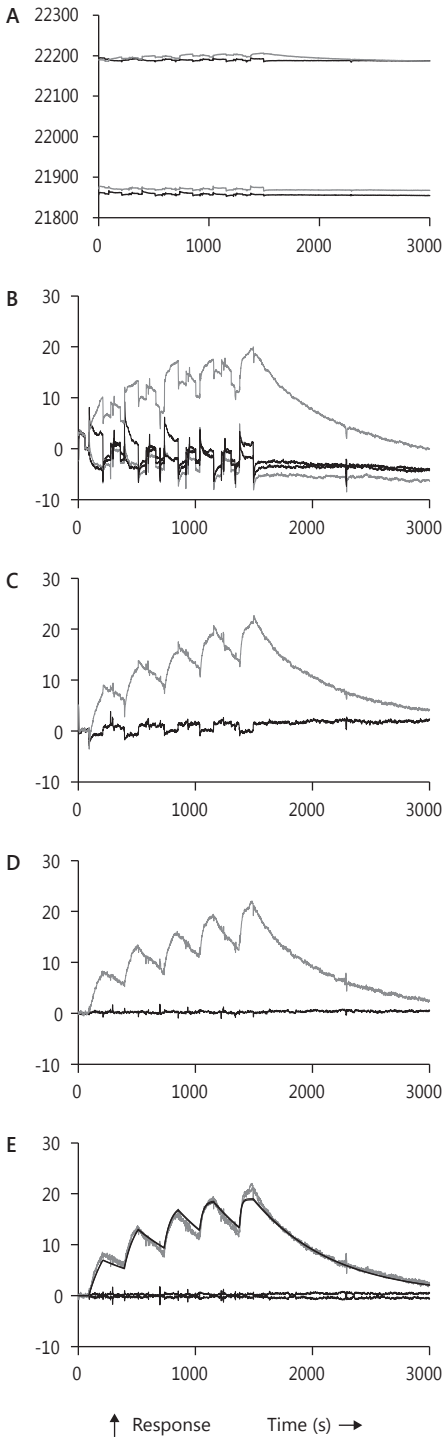


Figure 2.7: Data processing. A) Raw data from receptor surface (top) and reference surface (bottom) of buffer injections (black) and analyte injections (gray). B) data is zeroed to the y-axis prior to the first injection. C) Data of the reference surface is subtracted. D) The average of multiple blank injections is subtracted. E) Replicate injections are overlaid and fitted to a titration kinetics model (smooth black line). Surface density: 678 RU, flow: 30 µl min^{-1}, analyte concentrations: 6.25, 12.5, 25, 50 and 100 nM. Fitting yielded for k_a: 6.6 × 10^5 M^{-1} s^{-1}, for k_d: 1.5 × 10^{-3} s^{-1}, for K_D: 2.2 nM.

Concluding remarks

Obtaining high quality kinetic data is usually not straightforward, but it is worth the effort because such data is often easy to analyze and generally follows a 1:1 binding behavior. A proper fit to the data, presented in a clear figure is more than a satisfying reward; it will provide the readers of your paper with confidence in the data and shows them that you know what you were doing. Unfortunately this understanding has not been fully recognized, as becomes clear from the amount of poor quality data that still is being published. After reading this chapter, we hope you are encouraged to only generate and present high-quality SPR data.

Supplement – An example

The data shown in Figure S2.1 was obtained within the framework of the 2011 Global Label-free Interaction Benchmark study (Rebecca L. Rich & David G. Myszka, Biosensor Tools, LLC; Salt Lake City, USA). Benchmark studies serve the important purpose of educating users but also to gain information on the reliability and variability of the technique as a whole [236]. In this particular study, participants were provided with a biotinylated target molecule (27.3 kDa), an analyte (20 µM, 50.8 kDa), a streptavidin-coated sensor chip (Xantec, SAD 200M), and the task to analyse the interaction. To standardize the analysis, the temperature was fixed to 25°C and the running buffer defined (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.01% tween 20, 0.1 mg ml^{-1} BSA, pH 7.4).

During the experiment the flow was kept constant at 30 µl min^{-1}, while ligand was captured at a flow of at 5 µl min^{-1}. As initial test, the biotinylated target was captured to a density of 110 RU in fc 4. Because the analyte molecule is approximately two times heavier than the ligand the maximal response is expected to be around 205 RU, according to equation 2.4 (that assumes 1:1 binding).

Equation 2.4:
$$\text{Analyte binding capacity (RU)} = \frac{\text{MW analyte}}{\text{MW ligand}} \times \text{immobilized ligand (RU)}$$

A test injection of 100 nM analyte, however, only yielded a response of 13 RU; probably the ligand was not fully active anymore. In any case, this was considered too low to perform a full kinetic analysis, and hence the analyte was captured to higher densities in the remaining channels: 424 RU in fc 2 and 678 RU in fc 3; fc 1 was left empty to serve as reference channel. Based on the test injection it was estimated that a regeneration step would not be necessary, provided the dissociation phase was long enough. Therefore the titration kinetics approach was chosen to be suitable.

Five concentrations of either a 2 fold or 3 fold dilution series were injected for 2 min, followed by a 1 min dissociation phase. The last and highest concentration, 100 nM, was followed by a dissociation phase of 45 min, of which the last 20 min were excluded from the analysis. Each dilution series was injected in triplicate and

two buffer injection cycles were performed, one preceding each set of triplicates. With hindsight more buffer injections should have been performed to increase data quality, because the first of each triplicate does not overlay as nicely with the other two (Figure S2.1). Double referenced data was globally fitted using a titration kinetics model [126]. Although the fits are not perfect, the results show that different surface densities and dilution series yield comparable kinetics (Figure S2.1).

As it turns out, after returning the data, the biotinylated ligand molecule was actually a mix of two forms of the protein; both forms bind the analyte with distinct affinity and kinetics. By creating a heterogeneous surface, the initiators of the benchmark study set out to mimic complexity that may be encountered in regular experiments. During the analysis data should have been fitted with a model specific for a heterogeneous surface. Although such a model is available in the Biacore 3000 analysis software for a classical kinetics assay, it is not available for a titration kinetics assay. Therefore, and despite that the fits are not optimal, the fits of the 1:1 model are the best approximation of the kinetics and affinity of the interaction.

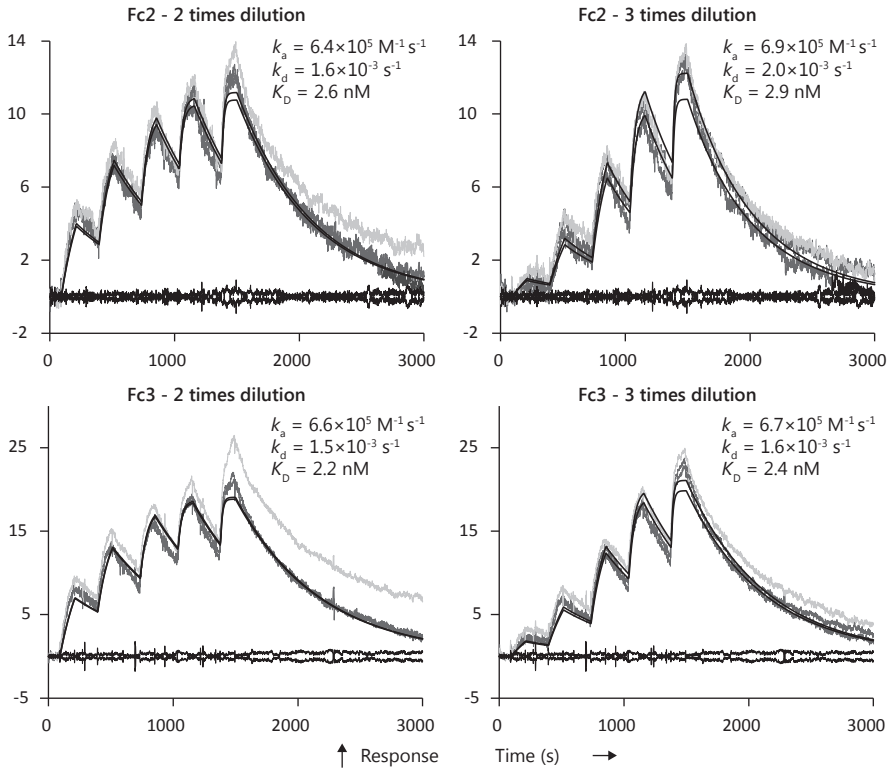


Figure S2.1: Results of the titration kinetics series. Two different surface densities and two different dilution series were used. Data points used for fitting are shown in dark gray, baseline responses are shown in black. Fits to the data are indicated by the smooth black line. The first injection series of each dilution series that was not included in the fitting is shown in light gray.

Chapter 3

Kinetic and stoichiometric characterization of streptavidin-binding aptamers

Vincent J. B. Ruigrok, Esther van Duijn, Arjan Barendregt, Kevin Dyer, John A. Tainer, Regina Stoltenburg, Beate Strehlitz, Mark Levisson, Hauke Smidt, and John van der Oost

ChemBioChem, 2012. 13(6): p. 829-836.

Abstract

Aptamers are oligonucleotide ligands that are selected for high-affinity binding to molecular targets. Only limited knowledge relating to relations between structural and kinetic properties that define aptamer-target interactions is available. To this end, streptavidin-binding aptamers were isolated and characterized by distinct analytical techniques. Binding kinetics of five broadly similar aptamers were determined by surface plasmon resonance (SPR); affinities ranged from 35–375 nM with large differences in association and dissociation rates. Native mass spectrometry showed that streptavidin can accommodate up to two aptamer units. In a 3D model of one aptamer, conserved regions are exposed, strongly suggesting that they directly interact with the biotin-binding pockets of streptavidin. Mutational studies confirmed both conserved regions to be crucial for binding. An important result is the observation that the most abundant aptamer in our selections is not the tightest binder, emphasizing the importance of having insight into the kinetics of complex formation. To find the tightest binder it might be better to perform fewer selection rounds and to focus on post-selection characterization, through the use of complementary approaches as described in this study.

Introduction

Aptamers are oligonucleotides (RNA or single-stranded DNA) that serve as high-affinity binding ligands, selected to interact specifically with a target. Potential aptamer targets range from small organic molecules (e.g., ethanolamine [165] and acetylcholine [24]) to proteins and large protein complexes [18, 285], and even to cells [27, 281]. Research into aptamers has steadily increased after their initial development in 1990 [58, 285]. Although at present there are only a few commercial applications of aptamers, they certainly have beneficial properties for application as therapeutics [e.g., FDA-approved Pegaptanib (Macugen) [82]]. In addition, aptamers have potential in the fields of targeted drug-delivery systems, biological recognition tools in biosensors and imaging tools [245].

Aptamers that specifically bind predefined targets are generally selected *in vitro* by SELEX (systematic enrichment of ligands by exponential enrichment). The SELEX procedure consists of multiple selection rounds, starting with a large pool of synthetic oligonucleotides containing a variable region (10^{14}–10^{15} variants) flanked by constant regions. Rounds of selection are initiated by exposure of RNA or ssDNA oligonucleotides to an immobilized target molecule. Subsequently, non-binding oligonucleotide molecules are washed away, after which bound molecules are recovered, amplified by PCR or RT-PCR and made single-stranded again. Several selection procedures have been successfully applied over the years, including affinity chromatography [305], capillary electrophoresis [177] and enrichment through the use of target-coated magnetic beads [268]. After successful selection, the minimal binding sequence should be obtained, in order to increase specificity and to reduce synthesis costs. Approaches to identification and characterization of the minimal binding sequence can be straightforward, but only when a conserved nucleotide motif and folding pattern is enriched.

A relevant feature of an aptamer is its affinity for its target. The dissociation constant (K_D in M) is often measured by approaches that do not provide insight into the actual kinetics of binding. However, the kinetic parameters can reveal valuable information: the association rate constant (k_a in M^{-1} s^{-1}) is a measure of how rapidly a complex is formed, whereas the dissociation rate constant (k_d in s^{-1}) is a measure of how rapidly a complex falls apart. The dissociation constant can easily be calculated from the ratio of the dissociation and association rate constants ($K_D = k_d / k_a$); this illustrates how different kinetics can result in similar affinities. Knowledge of underlying kinetic parameters is therefore relevant for better understanding of the molecular basis of complex formation and stability. One of the methods used to determine kinetic parameters (k_a and k_d) is surface plasmon resonance (SPR), an advanced method for following binding interactions in real time [49].

Streptavidin, a 60 kDa homo-tetrameric protein originally isolated from *Streptomyces avidinii*, is used extensively in molecular biology, because the

interaction between streptavidin and its natural ligand biotin is one of the strongest non-covalent molecular interactions known (K_D ~10^{-14} M). In the absence of biotin, a loop close to the binding pocket is open; however, upon biotin binding the loop moves to cover the pocket. This conformational change explains the exceptionally high affinity –the result of a very low k_d value– for biotin [148]. In two recent studies, streptavidin has been used as target for aptamer selection. Remarkably, the predicted secondary structures of the enriched binding motifs were quite similar, although only a few nucleotides appeared to be conserved [14, 268]. From equilibrium analyses and competition experiments, dissociation constants ranging from 40 to 85 nM were reported, although no data on the kinetics are available. In addition, biotin was shown to compete with streptavidin-binding aptamers, suggesting that the aptamers bind in or near the biotin-binding pocket [14]. Nevertheless, the stoichiometry and the location of binding are not known.

Here we set out to test whether or not multiple independent selection experiments would result in similar aptamers, in terms of sequence and structure as well as kinetic properties. In addition, we wanted to obtain a thorough understanding of the kinetics underlying aptamer selection, ideally resulting in aptamers with improved binding affinities. Streptavidin was chosen as target because streptavidin-binding aptamers have already been described, because detailed knowledge on its structure is available, and because its stability makes it relatively easy to perform SPR experiments. Seven rounds of selection resulted in streptavidin-binding oligonucleotides with a conserved secondary structure. Five different aptamer families could be distinguished, and SPR analysis showed that, despite their similarity, they differ greatly in their affinities (35–375 nM) and their kinetic behavior. In addition, the stoichiometries of the complexes were determined by native mass spectrometry, and small angle X-ray scattering (SAXS) data were used to generate a structural model of the aptamers.

Results

Streptavidin-binding aptamers

Through the use of the experimental approaches described later in this chapter we aimed to obtain ssDNA aptamers that would bind to streptavidin. After seven rounds of selection, the enriched pool of aptamers was cloned. In total, 91 clones were sequenced and used for subsequent analysis. Assembly of individual sequences yielded 12 clusters that each contained two to 13 fully identical sequences, whereas 19 sequences occurred only once, resulting in a total of 31 unique sequences (Figure S3.1 in the Supplementary information).

Structural predictions of all unique sequences (clusters and singlets) were made using mfold (a selection is depicted in Figure S3.2). In 26 out of the 31 sequences a similar structure of a stem-bulge-stem-loop was predicted (Figure 3.1); in the remaining five sequences (two clusters representing 16 clones, and three singlets) no

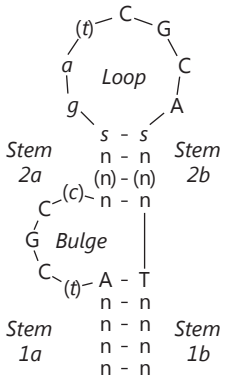


Figure 3.1: Secondary structure of the predicted stem-bulge-stem-loop sequences. Nucleotides in capitals are conserved in the majority of sequences, nucleotides in lower case italics are conserved within a family (or families), and nucleotides between parentheses are absent in some families. n represents one of four nucleotides, and s represents C or G.


```
                 Stem1a    Bulge Stem2a    Loop     Stem2b Stem1b
StrepApt1
Ctg_03(5)     ------GAA---CGCGT---GTG---TT--GCT---CAC----TTC
A03           ATTGTTCCA---CGCGT---GTG---TT--GCT---CAC----TGGCACAAT
A12           ------CCA---CGC----GTAA---TT--GGG---TTAC---TGG        10%
D07           --CTTTTCA---CGC----ATGA---TT--GGG---TCAT---TGAAAAG
G11           ----GTCGA---CGC----GTTG---TT--GGG---CAAC---TCGAC
StrepApt2
Ctg_02(12)    -----GGTA--TCGCT---ACCT---GA-CGCA---AGGT---TACC
Ctg_05(13)    -----CGAA---CGC-----ATC--TGA-CGCAG--GAT----TTCG
Ctg_06(4)     ------CAA--TTGCC---AACC---GA-CGCA---GGTT---TTG
Ctg_07(11)    ---ATTCAA--TTGCC---ACGT---GA-CGCA---ACGT---TTGAAT    53%
Ctg_09(4)     ---AAGGTG--CGGCT---ATTC---GA-CGCA---GAAT---CAGCTT
Ctg_11(2)     ------CAA---TGC----CCCT---GT-CGCA---AGGG---TTG
A07           ----CATCA--TCGCC---ACCC---GA-CGCA---GGGT---TGATG
F05           ----CGATT---CGC----GTAC---GT-CGCA---GTAC---AATCG
StrepApt3
Ctg_10(2)     ------CTA---CGC----GAGT---GC-CGCA---ACTC---TAG
Ctg_12(10)    ---AGATCA---CGC-----GAC---GC-TGCA---GTC----TGATCT
A05           ATTGTAACA---CGC----GTGC---GG-CGCA---GCAC---TGTACAAT
C04           ---GTACGG--TCGCC--CCCAT---GC-CGCA---ATGGG--CCGTAC    19%
D01           -TTACGATT---CGC----GTAC---GC-TGCA---GTAC---AATCGTAA
E02           -----ACGA--CCGCT---ATAC---GC-CGCA---GTAT---TCGT
F04           TATTACTAA---TGC----CCAC---GG-CGCA---GTGG---TTAGTAAG
StrepApt4
Ctg_01(2)     --ACATTAA---CGC-----CCG---TATTGCT---CGG----TTAATGT
G04           -----AGAA---CGC-----CCG---TATTGCT---CGG----TTCT      6%
G05           ----GTGAA---CGC-----GGG---TGTTGCT---CCC----TTCAC
G07           ------CCG---CGC-----TCG---TATTGCT---CGA----CGG
StrepApt5
Ctg_04(5)     ----AAAAG-GACGC-----ACC---GATCGCA---GGT-TA-CTTTT
Ctg_08(2)     ---TTTCCA---CGC-----ACC---GATCGCA---GGT----TGGAAA
B05           -CAATTGAA---CGC-----tCC---GATTGCA---GGA----TTCAGTTG
E04           ------TTA---CGC-----ACC---GATCGCA---GGT----TAA       12%
E07           -----TGTA---CGC-----ACC---GATCGCA---GGT----TACA
G09           -----ACCA---CGC-----ATC---GATCGCA---GAT----TGGT
H07           -----AGTT---CGC-----ACC---GATCGCA---GGT----AACT
```

Figure 3.2: Alignment of stem-bulge-stem-loop sequences. Nucleotides conserved in the majority of sequences and those conserved in families are highlighted. Ctg is a contig based on the number of sequences indicated in parentheses. Percentages indicate the occurrence of a certain family within the clone library.

such structure was predicted. After manual sequence inspection, however, regions that could form stem-bulge-stem-loop structures were found in these five sequences as well. Submission of these sequence fragments to mfold confirmed this prediction. As expected, the locations of the structure motifs are not fixed within the DNA sequences; on some occasions the primers, flanking the random regions, seem to be involved in the structures (Figure S3.1).

In order to gain insight into the conservation of nucleotide sequences and structures, the sequences predicted to form the stem-bulge-stem-loop structures were manually aligned according to their positions within the predicted structures (Figure 3.2). Despite the overall structural similarity some small variations could be observed: loop sizes vary between five and seven nucleotides, the bulges contain either three or five nucleotides, and stem 2 has three base pairs when the loops contain seven nucleotides and four base pairs in cases of smaller loop sizes. In particular, the nucleotides in the bulges (CGC) and loops (CGCA) are well conserved, whereas the actual compositions of the stems do not seem to be important as long as proper Watson-Crick base pairing is maintained.

From the alignment and the small differences between sequences it was possible to distinguish five aptamer families that occur at different frequencies in the library of selected clones. A shortened aptamer representing each family –StrepApt1 to StrepApt5– was designed for subsequent experiments while maintaining the specific features of each family (Figure 3.3).

Aptamer binding kinetics

Shortened aptamers –StrepApt1 to StrepApt5– as well as St-2-1 described by Bing *et al.* [14], were used for the determination of aptamer binding kinetics by SPR. A titration-kinetics approach was used: in one run increasing aptamer concentrations

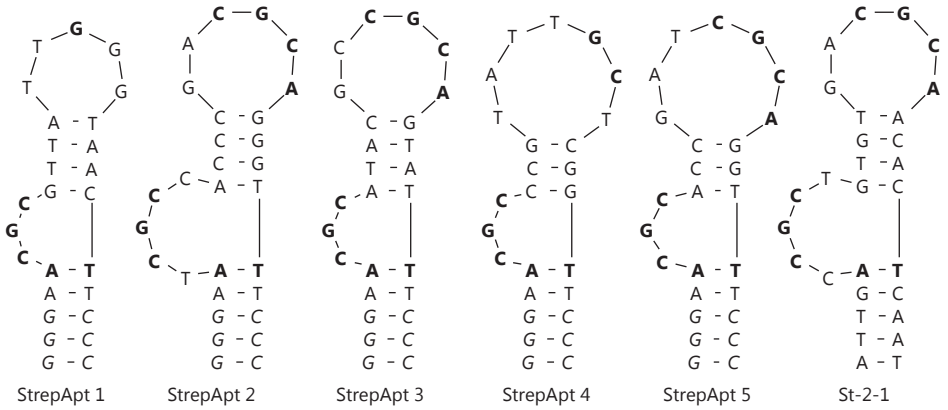


Figure 3.3: Shortened aptamers representing aptamer families. Nucleotides in bold are those conserved in the majority of sequences. Nucleotides in italics are not based on specific features in the alignment, but instead are added to assure stable conformations.

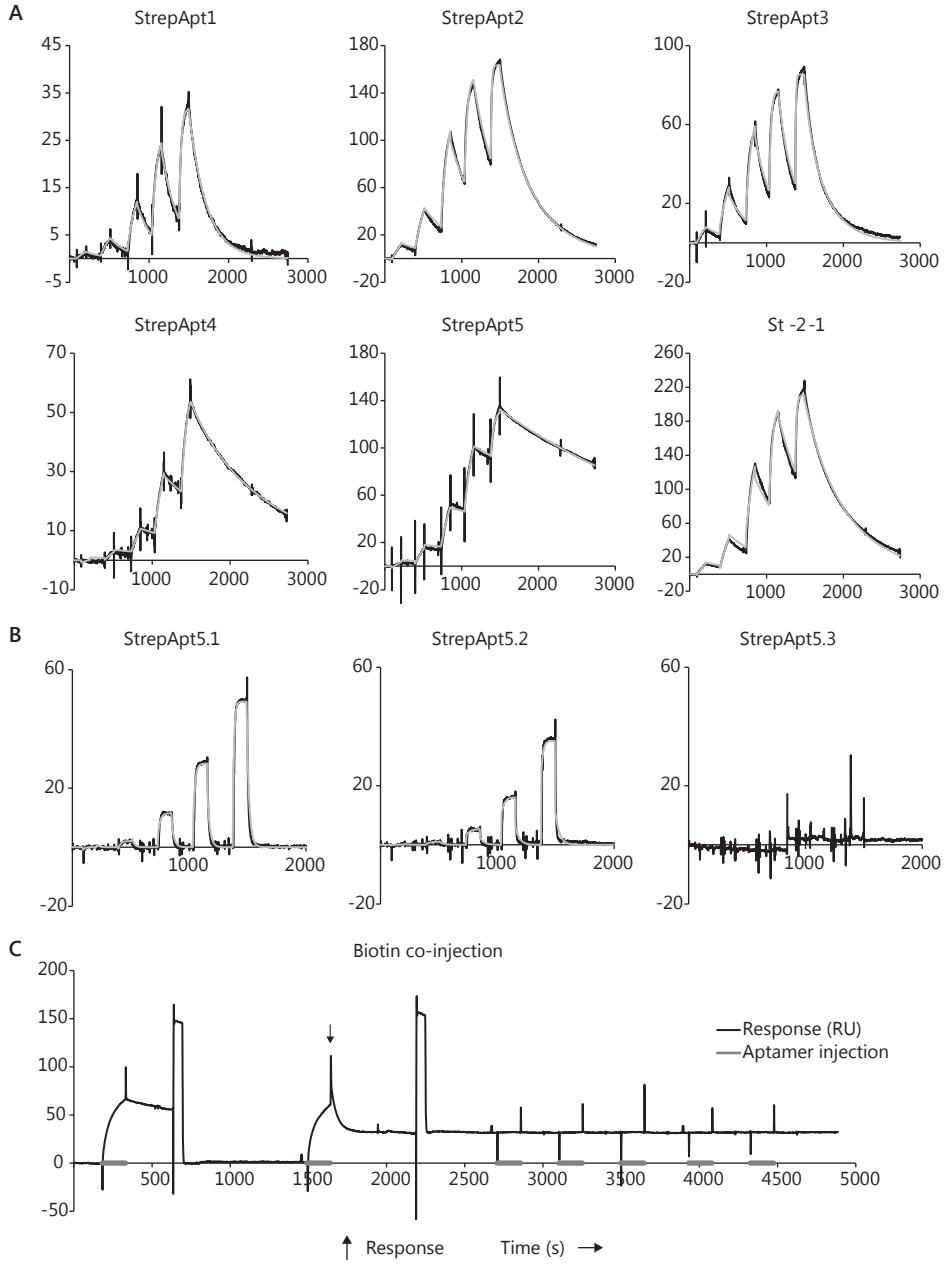


Figure 3.4: SPR data of streptavidin-aptamer interactions. A) Titration kinetics of shortened aptamers, different shapes indicate a different kinetic behavior (see table 3.1 for details). B) mutated variants StrepApt5.1 and 5.2 show reduced binding affinity, while StrepApt5.3 does not show any binding at all. C) Addition of biotin, indicated by an arrow, results in a more rapid aptamer dissociation. Subsequent aptamer injections do not result in binding, suggesting that the binding sites are blocked. Double referenced are shown in black; fits according to a simple 1:1 binding model are depicted in gray (A and B).

were sequentially injected over a streptavidin-coated surface. Double referenced data were used for fitting according to a simple 1:1 binding interaction model designed for fitting titration kinetics data [126]. Experimental data and corresponding fits are shown in Figure 3.4A, whereas rate constants (k_a, k_d) and dissociation constants (K_D) derived from fitted data are presented in Table 3.1.

Although the differences in sequences and predicted structures are small, the binding kinetics of the five streptavidin aptamer types varied dramatically. In direct relation to the differences in their kinetics, the dissociation constants (K_D) range from 35 to 375 nM (Table 3.1). The dissociation constant found for St-2-1 (79±2 nM) agrees well with data reported earlier (40±18 nM) [14]. It should be noted that association and dissociation rates for this aptamer were not reported in that study because of limitations of the analytical approach used. From the association rates (k_a) of the five shortened aptamers designed in this study it was possible to distinguish two groups: those that form complexes rapidly (StrepApt2 and StrepApt3; Table 3.1) and those that form complexes slowly (StrepApt1, StrepApt4 and StrepApt5). From the dissociation rates (k_d) it was possible to distinguish two other groups: those that dissociate rapidly (StrepApt1, StrepAtp2 and StrepApt3) and those that dissociate slowly (StrepApt4 and StrepApt5).

StrepApt4 and StrepApt5 (low values of k_a and k_d) represent only minor fractions of sequences in the clone library, which is dominated by StrepApt2, representing 53% of all sequences. Despite being the most abundant sequence, StrepApt2 does not have the highest affinity (77 nM). Rather, the highest-affinity aptamer (StrepApt5, 35 nM) represents only 12% of the sequences in the clone library. It is interesting to note here that the affinity and kinetic behavior of St-2-1 are similar to those of StrepApt2 (Table 3.1).

Table 3.1: Relative occurrence of short the clone library. Together with kinetic parameters of aptamer-streptavidin binding as determined by SPR analysis. Highest association rates and slowest dissociation rates are highlighted in bold.

	Occurrence in clone library (%)	K_D (nM)	k_a ($\times 10^4$ M^{-1} s^{-1})	k_d ($\times 10^{-3}$ s^{-1})
StrepApt1	10	375±50	1.37±0.05	5.1±0.7
StrepApt2	53	77±3	**3.98±0.09**	3.07±0.09
StrepApt3	19	105±4	**4.8±0.8**	5.0±0.8
StrepApt4	6	181±7	0.73±0.07	**1.32±0.08**
StrepApt5	12	35±1	1.04±0.03	**0.36±0.01**
St-2-1	-	79±2	3.2±0.3	2.5±0.2

Influence of mutations and biotin binding

Two conserved regions, sharing a common CGC motif, were found in the majority of the shortened aptamers (Figure 3.3). In order to determine the importance of

the conserved guanines for streptavidin binding, they were replaced by thymines in StrepApt5, our best binder. Replacement of one guanine resulted in a strongly reduced affinity in the micromolar range (~2 µM; Figure 3.4B). Although we performed only one injection series for these mutants the data were still fitted for the purpose of obtaining a rough estimate of the K_D value. Replacement of both guanines by thymines completely abolished binding, indicating that these residues are essential for binding in all aptamer families.

SPR was furthermore used for competition experiments. Biotin is the natural ligand bound by streptavidin, with an exceptionally high affinity (K_D ~10^{-14} M) [148]. In order to assess the impact of biotin on aptamer binding, the streptavidin-coated chip was first loaded with StrepApt5, the aptamer with the highest affinity, followed by normal dissociation (Figure 3.4C). After regeneration the chip was reloaded with aptamer, directly followed by injection of biotin (by use of the coinject command). The faster dissociation and increased baseline, even after regeneration, revealed that biotin out-competes StrepApt5. In addition, subsequent aptamer injections did not result in any binding.

Stoichiometries of the complexes

The binding stoichiometries of the various streptavidin-aptamer complexes were determined by native mass spectrometry. Native mass spectrometry is a powerful method for study of non-covalent complexes, the exact mass analysis often being sufficient to determine the stoichiometry of a complex unambiguously. Streptavidin was found to bind a maximum of two aptamer units simultaneously, regardless of the aptamer used (Figure 3.5). In addition to identifying binding stoichiometries, native mass spectrometry can also be applied to monitor the binding strengths of aptamers to streptavidin. The relative abundances of free streptavidin, of streptavidin bound to one aptamer and of streptavidin bound to two aptamers confirmed that aptamer 5 has the highest affinity for streptavidin (Figure 3.5B and Table S3.1).

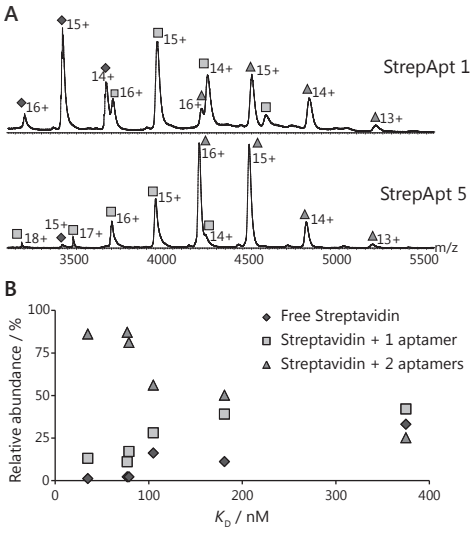


Figure 3.5: Native mass spectrometry data. A) Spectra of the lowest-affinity binder and of the highest-affinity binder. B) Relative abundances of free streptavidin, streptavidin bound to one aptamer and streptavidin bound to two aptamers. Diamonds show data corresponding to free streptavidin, squares data corresponding to streptavidin+1 aptamer, and triangles data corresponding to streptavidin+2 aptamers.

Structural model of StrepApt2 and the complex

Modelling of the 3D structure of StrepApt2 was performed with use of the MC-Fold and MC-Sym pipeline to provide insight into the structural basis of aptamer binding to the streptavidin tetramer. In total, 57 3D structures were predicted for StrepApt2, from its sequence and the 2D structure predicted by mfold. In order to identify the best prediction, SAXS data for the aptamer were compared with theoretical scattering of the 57 predicted models. The structure (Figure 3.6) with the lowest Chi^2 value (5.32, Figure S3.3) was considered the best match with the experimental scattering. This structure indicates that the conserved nucleotides in the loop (CGCA) and bulge (CGC) are solvent-exposed and are therefore the most likely candidates to interact with streptavidin.

Figure 3.6: Predicted 3D model of StrepApt2. Prediction 28 of the MC-Fold and MC-Sym pipeline that best matches experimental scatter. Stems are in gray, bulge and loop in light gray. The two guanines that are essential for binding are depicted in black (and are indicated by arrows).

Discussion

DNA aptamers that bind to streptavidin, with affinities in the low to middle nanomolar range (35–375 nM), were enriched in seven sequential rounds of selection. Sequence analysis of 91 clones revealed 31 unique sequences at different relative abundances. Structural predictions, based on mfold, and sequence analysis show that all sequences contain stem-bulge-stem-loop structures. Interestingly, similar structures for streptavidin-binding aptamers had been reported earlier, although the aptamer sequences vary slightly, and nucleotides in the bulge (CGC) and loop (CGCA) regions seem to be conserved throughout various independent selection experiments [14]. Variation in loop composition and size (five to seven nucleotides), bulge size (either three or five nucleotides) and stem 2 length (three or four base pairs) further allowed five aptamer families to be distinguished. The compositions of the stems do not seem to be important as long as proper Watson-Crick base pairing is maintained. The length of stem 1, however, is important for stability of the aptamer, as has been shown previously by Bing and co-authors [14]. Shortened aptamers –StrepApt1 to StrepApt5– for each of the five types, based on the alignment, were designed, with the specific features of each family maintained. Single mutations of the conserved guanines in the loops and bulges resulted in reduced binding of the aptamers, whereas in the double mutant binding was abolished, indicating that these conserved guanines play an essential role in streptavidin binding.

We further determined the kinetic properties of the shortened aptamers StrepApt1

to StrepApt5 by use of SPR titration kinetics. The obtained results showed widely different dissociation constants (35–375 nM), but no obvious relation between affinity and abundance in the clone library was observed. The number of clones is limited and might not represent the complete range of aptamers present after seven selection rounds, but results of independent studies showed enrichment of similar aptamers. Kinetic parameters of representative aptamers indicate that the rapid binders (StrepAtp2 and StrepApt3, loop size 6 nt) dominate the clone library, together representing 72% of all sequences. In contrast, the slowly dissociating aptamers (StrepApt4 and StrepApt5, loop size 7 nt) are a minority, together representing 18% of the clone library. The kinetic parameters determined in this study show that slow dissociation seems to be related to a loop size of seven nucleotides, whereas rapid association seems to be associated with a loop size of six nucleotides. It is interesting to note here that the aptamer with the highest affinity was not the most abundant in our clone library. Because of its structural similarity we also included St-2-1, selected by Bing and co-authors [14], in our SPR analysis. The kinetic parameters of St-2-1, as well as its predicted folding motif, were most similar to those of StrepApt2. Two independent selection experiments hence resulted in similar binders, with rapid association rates; it is therefore tempting to speculate that the selection procedure is somewhat biased towards selection of rapid binders, rather than high-affinity binders. This is also observed in the extensive work by Win *et al.* [306] on codeine-binding RNA aptamers. These authors found several sequences to be present more than once in their clone library, whereas the five best binders, those with the lowest K_D values, occurred only once.

Consistently with the study of Bing *et al.*, we also observed competition between aptamer (StrepApt5) and biotin. Nevertheless, our data clearly showed aptamer binding to be completely suppressed when streptavidin is saturated with biotin, whereas Bing *et al.* still observed some binding in the presence of biotin [14]. Both studies, however, unequivocally suggest that the aptamer binds in, or close to, the biotin-binding pocket.

To gain insight in the stoichiometry of the streptavidin-aptamer complex we performed native mass spectrometry, which indicated that streptavidin can bind a maximum of two aptamer moieties simultaneously, regardless of the aptamer used and its affinity. This leaves the question, however, of whether one aptamer binds two biotin pockets at once, or if access to a single biotin pocket is blocked by a bound aptamer. In order to gain insight in the location of binding, and to provide an answer to the question of whether or not one aptamer binds two biotin pockets, structural predictions (57) of StrepApt2 were generated by use of the MC-Fold and MC-Sym pipeline. The conserved nucleotides in the loop and bulge are solvent exposed and are therefore the most likely candidates to interact with the loops that cover the two biotin pockets once biotin is bound (Figure S3.4). On the basis of this model it can be

hypothesized that one aptamer occupies two biotin pockets. This is supported by the presence of two conserved regions and by the fact that two simultaneous mutations, one in each of these conserved regions, are required to yield a non-binding aptamer. Our approach of obtaining a 3D model for an aptamer, supported by SAXS data, might also be helpful as a new strategy to afford insight into the binding sites of other aptamers.

Conclusions

All streptavidin-binding aptamers selected in this work have similar structures and conserved nucleotides, but still vary slightly at sequence level. Although the differences in sequence are small, the differences in binding kinetics are substantial. One streptavidin tetramer can accommodate up to two aptamers, and we propose that one aptamer occupies two biotin-binding pockets. It is interesting to note that the aptamer with the highest affinity is not the most abundant. Instead, selection appears to be driven by the binding kinetics. This could be the result of the presence of a limited number of binding sites that are quickly saturated by rapid binders when selection progresses through subsequent rounds. This finding indicates that it might be better either to increase the number of target binding sites (by using more beads, or less DNA) or to limit the number of selection rounds. Moreover, these findings indicate that one should focus on post-selection characterization of potential aptamers, by using complementary approaches introduced here, for example, in order to find the aptamer with the highest affinity.

Materials and methods

Oligonucleotides

Unmodified primers AP10 (ataccagcttattcaatt) [268] and AP30 (ctaactgattacgattgt) (R. Stoltenburg, personal communication), the random pool (ataccagcttattcaatt-n_{64}-acaatcgtaatcagttag, PAGE purified), primer AP60, equivalent to AP10 but containing 5′-fluorescein, and shortened aptamers (for sequences see Table 3.2) were obtained from Biolegio (Nijmegen, the Netherlands). Primer TER-AP30, equal to AP30 but containing 5′-poly-dA_{20}-hexa(ethylene glycol), was ordered from IBA (Göttingen, Germany).

SELEX procedure (FluMag-SELEX)

The selection procedure was as described previously [268]. Briefly, before each selection round ssDNA was heated to 95°C for 8 min, cooled on ice for 10 min and left at room temperature for at least 10 min. In the first round, 26 pmol of the random pool was added to streptavidin-coated beads (Dynabeads M-280, Invitrogen), and in subsequent rounds 250 µl of the ssDNA preparation (see below) was added. Approximately 1×10^8 beads were washed with selection buffer [NaCl (100 mM), $MgCl_2$ (2 mM), KCl (5 mM),

$CaCl_2$ (1 mM), Tris-HCl (pH 7.6, 20 mM)], and after addition of the ssDNA to the beads they were incubated at 25°C with mild shaking for 30 min. Unbound oligonucleotides were removed by washing three times with selection buffer. Subsequently, bound oligonucleotides were eluted by addition of selection buffer and incubation at 95°C for 5 min with mild shaking.

Table 3.2: Sequences of shortened and mutated aptamers used in this study. Uppercase letters indicate nucleotides that are conserved in the majority of sequences within a given family. G→T mutations in StrepApt5.1, StrepApt5.2 and StrepApt5.3 are given in bold letters and underlined. All sequences are shown in the 5'-to-3' direction. The sequence of St-2-1 was taken from [14].

	Sequence
StrepApt1	gggaACGCgttaTTGGGtaacTtccc
StrepApt2	gggaatCGCcaccCGACGCAgggtTtccc
StrepApt3	gggaACGCataCGCCGCAGtatTtccc
StrepApt4	gggaACGCccgTATTGCTcggTtccc
StrepApt5	gggaACGCaccGATCGCAggtTtccc
StrepApt5.1	gggaAC**<u>t</u>**CaccGATCGCAggtTtccc
StrepApt5.2	gggaACGCaccGATC**<u>t</u>**CAggtTtccc
StrepApt5.3	gggaAC**<u>t</u>**CaccGATC**<u>t</u>**CAggtTtccc
St-2-1	attgaccgctgtgtgacgcaacactcaat

Eluted oligonucleotides were PCR-amplified in 15 parallel PCR reactions, each mixture containing $MgCl_2$ (1.9 mM), dNTPs (0.2 mM each), primer (AP60 and TER-AP30, 1 µM of each), 1×buffer B2, and HOT FIREPol (Solis BioDyne, 4 units) in a total volume of 100 µl. The PCR program was as follows: 15 min at 95°C, 30 cycles of 30 s at 95°C, 30 s at 52°C, 1 min at 72°C, followed by 7 min at 72°C after the last cycle. Electrophoresis on a 12% polyacrylamide (PAA) gel was used to confirm successful amplification of a DNA fragment of the correct size.

From the second selection round onwards, fluorescence in the starting sample, the non-bound DNA fraction, the wash fractions and the eluted DNA was measured in 96-well microtiter plates (Greiner bio-one) with a SpectraMax M2 instrument (Molecular Devices).

ssDNA preparation

Unpurified PCR products were pooled and concentrated by ethanol precipitation and resuspended in TE buffer [Tris (10 mM), EDTA (1 mM), pH 7.4, 100 µl]. Strands were separated by electrophoresis of the concentrated sample on a denaturing PAA gel, containing PAA (8%) and urea (7 M) in TBE buffer [Tris (10 mM), borate (90 mM), EDTA (2 mM)]. Fluorescein-labelled DNA was visualized with a Safe imager (Invitrogen) and excised from the gel. The DNA was eluted from the gel slice by mashing and subsequent incubation in 600 µl EDTA (2 mM) and sodium acetate (pH 7.8, 300 mM) at 80°C with mild shaking for 2 h. After ethanol precipitation the DNA was resuspended in selection buffer (300 µl).

Cloning, sequencing and structure analysis

The DNA eluted in SELEX round 7 was amplified, under the conditions described

above, by using unmodified primers AP10 and AP30 and ligated into pGEM-T (Promega). *Escherichia coli* XL1-Blue cells (Stratagene) were transformed with the vector constructs, 96 colonies were subsequently randomly picked, and inserts were sequenced (GATC, Konstanz, Germany). The DNA folding form of mfold [324] was used for secondary structure prediction with use of the default settings except for temperature (25°C) and ionic conditions [Na^+ (100 mM), Mg^{2+} (2 mM)].

Native mass spectrometry

High-resolution mass spectra of streptavidin complexes with StrepApt1 to StrepApt5 and St-2-1 were recorded with an LCT instrument (Waters, Manchester, UK) in positive ion mode. Samples were loaded into gold-plated borosilicate capillaries made in-house [with a Sutter P-97 puller (Sutter Instruments Co., Novato, CA, USA) and an Edwards Scancoat six sputter-coater (Edwards Laboratories, Milpitas, CA, USA)] for analysis with an LCT 1 mass spectrometer (Waters Corp., Milford, MA, USA), adjusted for optimal performance in high-mass detection [278]. A capillary voltage of 1300 V was used, with a sampling cone voltage of 90 V. The source backing pressure was elevated in order to promote collisional cooling to approximately 6 mbar. All samples were sprayed from ammonium acetate (pH 7.6, 125 mM). Calibration of the instrument was carried out with a cesium iodide solution (25 mg ml^{-1}). Processing of the acquired spectra was performed with MassLynx 4.1 software (Waters Corp., Milford, MA, USA). Minimal smoothing was used, after which the spectra were centered. The mass of the species was calculated with use of each charge state in a series. The corresponding intensities of each charge state were assigned by MassLynx and summed. This approach allows the relative quantification of all species in a sample.

Surface plasmon resonance

In a Biacore 3000 system (BIACORE, Uppsala, Sweden), streptavidin (recombinant produced in *E. coli*, SIGMA), dissolved in sodium acetate (pH 4.0, 10 mM, GE Healthcare), was immobilized on a CM5 chip at 25°C (to approximately 2000 RU). The "aim-for-immobilized-level" wizard was used, together with reagents provided in the amine coupling kit (GE Healthcare). A reference channel was prepared by activating and subsequent blocking of the surface by the same procedure. The shortened aptamers StrepApt1 to StrepApt5.3 and St-2-1 were dissolved in selection buffer at concentrations of 100 µM and subsequently diluted in selection buffer to concentrations of 2.50, 0.83, 0.28, 0.09 and 0.03 µM. Titration kinetics were performed by sequential injections of each concentration (60 µl, starting with the lowest concentration) at 30 µl min^{-1}, followed by 60 s dissociation. The fifth injection was followed by 20 min dissociation and subsequent regeneration with glycine (pH 3.0, 10 mM, 10 µl) at 10 µl min^{-1}. Titration kinetics reduce assay times as well as the number of chip regenerations required. Prior to aptamer injections a series of blank

injections, with only selection buffer, was run to allow for double referencing of the data. Selection buffer was used as running buffer. Each aptamer was injected three times and double referenced data were used for fitting according to a 1:1 binding model as described by Karlsson *et al.* [126]. The concentration series of StrepApt5.1 to StrepApt5.3, which are mutated variants of StrepApt5, were injected only once.

Small-angle X-ray scattering and 3D modelling

StrepApt2 was dissolved in selection buffer to concentrations of 4.2, 2.9 and 1.5 μg μl^{-1}. SAXS data were collected at the SIBYLS beamline at 10°C, by procedures described previously [111]. Briefly, the three sample concentrations and a buffer blank were measured in order (buffer, low, middle, and high concentration, and again buffer) at exposures of 0.5, 1.0 and 6.0 s. Scattering curves were merged with use of PRIMUS [142] in the ATSAS program suite [141]. The MC-Fold and MC-Sym pipeline was used to obtain predictions of the 3D structure of StrepApt2 [215]. Theoretical scattering of these predictions was calculated and compared with experimentally observed scattering of StrepApt2 with use of the FoXS web-server [252].

Acknowledgements

This work was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research [ALW-GO-PL/08-08]. We thank Willem Haasnoot (RIKILT Institute of Food Safety, Wageningen) for help with initial SPR experiments.

Supplementary information

Table S3.1: Numbers used to create figure 3.5B. Affinity constants (K_D) and relative abundances of free streptavidin, streptavidin+1 aptamer and streptavidin+2 aptamers, as measured by native mass spectrometry. At lower affinities more free streptavidin is present, whereas more Streptavidin+2 aptamers is present at high affinities.

	K_D (nM)	Free Streptavidin	Streptavidin +1 aptamer	Streptavidin +2 aptamers
StrepApt1	375	33	42	25
StrepApt2	77	2	11	87
StrepApt3	105	16	28	56
StrepApt4	181	11	39	50
StrepApt5	35	1	13	86
St-2-1	79	2	17	81

```
                Primer      |---------------------- Random region --------------------------|    Primer

Contig_01_(2)   ATACCAGCTTATTCAATT CCGCGCGAACACATTAACGCCCGTATTGCTCGGTTAATGTAGGGGTCTGCCTGAACTCCCCGGT ACAATCGTAATCAGTTAG
Contig_03_(5)   ATACCAGCTTATTCAATT GCCACACATCAATTTGTTGTGAGGTTGGTACCGTTATTGTTGAACGCGTGTGTTGCTCACTTCC ACAATCGTAATCAGTTAG
Contig_04_(5)   ATACCAGCTTATTCAATT CCAACCCCTCCAAAAGGACGCACCGATCGCAGGTTACTTTTTGAGGCTGTCCGAACCTCGTGCT ACAATCGTAATCAGTTAG
Contig_05_(13)  ATACCAGCTTATTCAATT CACGGACAGTAGTGTTGGGGTCGAACGCATCTGACGCAGGATTTCGTACCACCGTCCGCACTGT ACAATCGTAATCAGTTAG
Contig_07_(11)  ATACCAGCTTATTCAATT GCCACGTGACGCAACGTTTGAATAGGTATGTCGGCTGTCAGAATTTCATGCGAATATTGGTTGT ACAATCGTAATCAGTTAG
Contig_08_(2)   ATACCAGCTTATTCAATT CAGCAGCATTAGAGTTTTCCACGCACCGATCGCAGGTTGGAAATCTTCTGCGTTAGTTTTCCCC ACAATCGTAATCAGTTAG
Contig_09_(4)   ATACCAGCTTATTCAATT ACAAAAGGTGCCGCTATTCGACGCAGAATCAGCTTCTTGTTAATAATTTGACCTTATGCCCCGT ACAATCGTAATCAGTTAG
Contig_10_(2)   ATACCAGCTTATTCAATT GCATGCCCCGATATATCCTACGCGAGTGCCGCAACTCTAGTATATCGATGACTGCACTACGTGGT ACAATCGTAATCAGTTAG
Contig_11_(2)   ATACCAGCTTATTCAATT ACGGACCATGTATTTGTGATGAGGATGTCTCTGTTATTGTCAATGCCCCTGTCGCAAGGGTTGT ACAATCGTAATCAGTTAG
Contig_12_(10)  ATACCAGCTTATTCAATT CCCCGACACTAGATCACGCGACGCTGCAGTCTGATCTTGTTCGCGTAATGGACCGGCCGTTGT   ACAATCGTAATCAGTTAG
A03-T7[1]       ATACCAGCTTATTCAATT CGTCAGGGTAATAGCTTGTACGCTGCAGTACTGTATTGTTCCACGCGTGTGTTGCTCACTGGC  ACAATCGTAATCAGTTAG
A05-T7[2]       ATACCAGCTTATTCAATT CCAGGCCGCTTCCTTTTGCTTGCACTGTGAACGGTATTGTAACACGCGTGCGGCGCAGCACTGT ACAATCGTAATCAGTTAG
A07-T7[3]       ATACCAGCTTATTCAATT ACATGCGGCTCATCATCGCCACCCGACGCAGGGTTGATGAACGCAGATCTGCGAACGTTCCCGT ACAATCGTAATCAGTTAG
A12-T7[4]       ATACCAGCTTATTCAATT CCGGCAGGTTCGGGAAGTCCACGCGTAATTGGGTTACTGGTCTTGCCTTTTGCCGGGACACTCT ACAATCGTAATCAGTTAG
B05-T7[5]       ATACCAGCTTATTCAATT GAACGCtCCCgATTGCAGGATTCAGTTGATACGGTTATGATTTAAAAATTTAGCCGACCCTCTGC ACAATCGTAATCAGTTAG
C04-T7[6]       ATACCAGCTTATTCAATT GCCCATCaaTGGTATATGATGCGCTTCCCCGGGGTACGGTCGCCCCCATGCCGCAATGGGCCGT AC ACAATCGTAATCAGTTAG
D01-T7[7]       ATACCAGCTTATTCAATT GCACGACATACGATTTGAATTACGATCATTGGCTTATTGTTTTACGATTCGCGTACGCTGCAGT ACAATCGTAATCAGTTAG
D07-T7[8]       ATACCAGCTTATTCAATT CACCGCGGGTTATCCTTTCACGCATGATTGGGTCATTGAAAAGGGTGAGGGGTGCTCAACGGT ACAATCGTAATCAGTTAG
E02-T7[9]       ATACCAGCTTATTCAATT CCAAAGTCTTCACGAAAGGACGACCGCTATACGCCGCAGTATTCGTTTCCTGTGATATCCTGT ACAATCGTAATCAGTTAG
E04-T7[10]      ATACCAGCTTATTCAATTT ACGCACCGATCGCAGGTTAAGTGAATAGTTCCGTTGTCACTCATGCTTCATGTTGctCCCCGT ACAATCGTAATCAGTTAG
F04-T7[12]      ATACCAGCTTATTCAATT CCGCGATGATTATTACTAATGCCCACGGCGCAGTGGTTAGTAAGCGTATGCAATcGACGGCGGT ACAATCGTAATCAGTTAG
G04-T7[14]      ATACCAGCTTATTCAATT CGGCAACACTACTGTTGAATTGACTTCCCTCGATTATTAGAACGCCCGTATTGCTCGGTTCTGG ACAATCGTAATCAGTTAG
G05-T7[15]      ATACCAGCTTATTCAATT CCCAAACCAATGTTTTTGGCAGCCCTGTGAACGCGGGTGTTGCTCCCTTCACGTTATGCCATGT ACAATCGTAATCAGTTAG
G07-T7[16]      ATACCAGCTTATTCAATT CCCGGTCCTGAGTTGTCCATAACTTGACATCGGTTATTGTTCCGCGCTCGTATTGCTCGACGGT ACAATCGTAATCAGTTAG
G09-T7[17]      ATACCAGCTTATTCAATT ACCACATGTAAACGGTTACCACGCATCGATCGCAGATTGGTTGCCGTTTGCCCTACTGCCCCGT ACAATCGTAATCAGTTAG
G11-T7[18]      ATACCAGCTTATTCAATT CCAGCCCGGCATATTTTACTTGGCTCAGTCGACGCGTTGTTGGGCAACTCGACCTCCGTACGCT ACAATCGTAATCAGTTAG

Contig_02_(12)  ATACCAGCTTATTCAATT CGACGACATGCGGTATCGCTACCTGACGCAAGGTTACCCTAGGCGATTAATATTTCCACGTTGT ACAATCGTAATCAGTTAG
Contig_06_(4)   ATACCAGCTTATTCAATT GCCAACCGACGCAGGTTTTGGGAAGCCCCACCATCATTTTAACTCCCGTATGTTAATCCACGCT ACAATCGTAATCAGTTAG
F05-T7[13]      ATACCAGCTTATTCAATT GCACACCAGCCACTTTTCCTGCGCATGTTTGACTTATCGGATCCGATTCGCGTACGTCGCAGT ACAATCGTAATCAGTTAG
H07-T7[19]      ATACCAGCTTATTCAATT ACGACCCGGCCAGAGTTAGTTCGCACCGATCGCAGGTAACTTACGATTCGTGTtgaGtcGCGGT ACAATCGTAATCAGTTAG
E07-T7[11]      ATACCAGCTTATTCAATT ACGGACCatTGGGTTAGTGTACGCACCGATCGCAGGTTATCAGGACCCATGGGCGGTTTTCTAT ACAATCGTAATCAGTTAG
```

Figure S3.1: Sequences of all 31 unique aptamers. Primer sequences are included. Occurrence of contig sequences in the clone library is indicated in parentheses. Location of conserved secondary structure is highlighted bold and underlined. Aptamers given in italics were initially not predicted when providing the full length sequence.

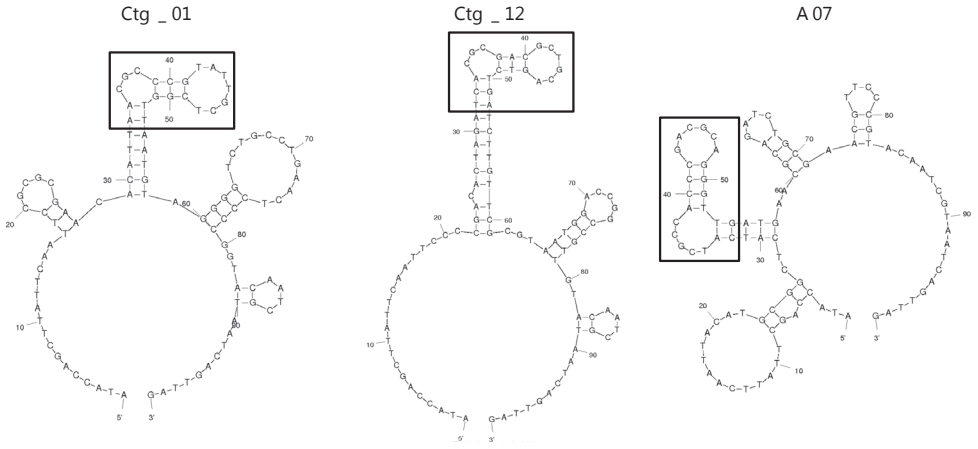


Figure S3.2: A selection of mfold predictions. Conserved binding domains are boxed.

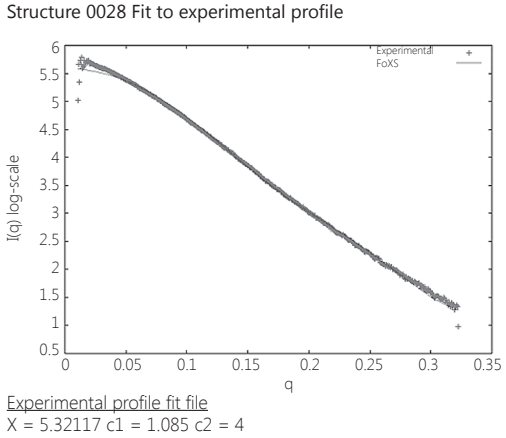


Figure S3.3: Comparison of theoretical scattering and actual scatter. Theoretical scatter of prediction 28, compared with actual scattering as determined by SAXS analysis. This prediction is considered the best match with experimental scattering.

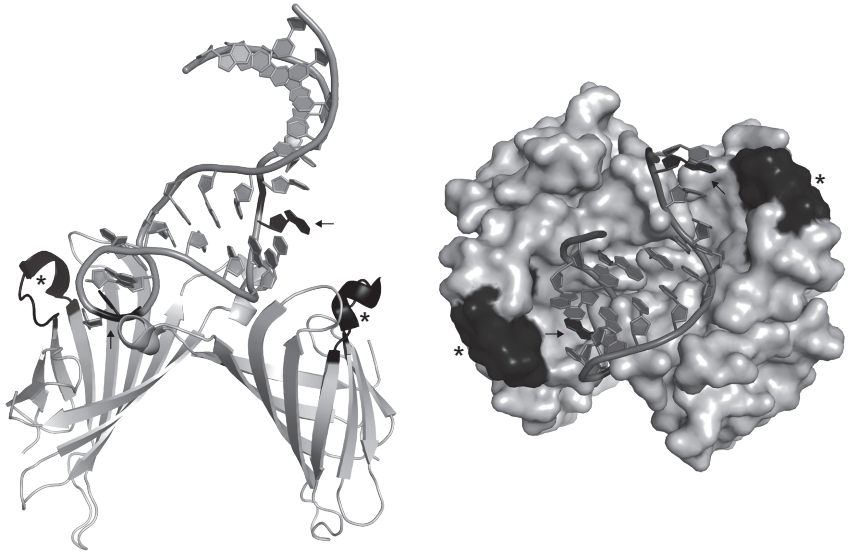


Figure S3.4: Impression of how the aptamer could bind to streptavidin. Conserved guanines (black arrow) probably interact with the loops that move over the biotin binding pocket (black, *). The aptamer model was 'fitted' manually and only serves to illustrate that the aptamer could physically fit in the space between the two loops. (PDB code: 1N9Y).

Chapter 4

Characterization of aptamer-protein complexes by X-ray crystallography and alternative approaches

Vincent J. B. Ruigrok [†], Mark Levisson [†], Johan Hekelaar, Hauke Smidt, Bauke W. Dijkstra and John van der Oost

[†] Contributed equally

International Journal of Molecular Sciences, 2012. 13(8): p. 10537-10552.

Abstract

Aptamers are oligonucleotide ligands, either RNA or ssDNA, selected for high-affinity binding to molecular targets, such as small organic molecules, proteins or whole microorganisms. While reports of new aptamers are numerous, characterization of their specific interaction is often restricted to the affinity of binding (K_D). Over the years, crystal structures of aptamer-protein complexes have only scarcely become available. Here we describe some relevant technical issues about the process of crystallizing aptamer-protein complexes and highlight some biochemical details on the molecular basis of selected aptamer-protein interactions. In addition, alternative experimental and computational approaches are discussed to study aptamer-protein interactions.

Introduction

Aptamers are single-stranded DNA (ssDNA) or RNA oligonucleotides selected to bind specifically to a predefined target. Since their initial development two decades ago [58, 285] the field of aptamer research has matured. During these years, numerous aptamers, recognizing a wide range of targets, have become available, such as, for example, those binding small organic molecules [176, 212, 308], peptides and proteins [257, 268], or even whole microorganisms [27, 153].

The systematic evolution of ligands by exponential enrichment (SELEX) is the most common method by which aptamers are enriched from large pools of randomized DNA or RNA (10^{14}–10^{15} variants), in an iterative process by applying several subsequent selection rounds. Selection is initiated by exposing the oligonucleotides to a target that is either coupled to a matrix (e.g., magnetic beads or column material) or already present on a surface (e.g., cell surface). Subsequently, non-binding oligonucleotide molecules are washed away, and the bound molecules are recovered, amplified by PCR or RT-PCR, and made single-stranded again. Several variations of the SELEX procedure have been successfully applied over the years [167].

In many cases, the minimal sequence required for efficient target binding is smaller than the length of the oligonucleotides that comprise the pool. In order to increase specificity and to reduce synthesis costs of the selected aptamers, the minimal binding sequence should be identified. Approaches to identify the minimal binding sequence can be straightforward when a conserved nucleotide motif or secondary structure is enriched [14, 247]; otherwise, identifying the minimal binding sequence can be more laborious [128, 160, 304].

Aptamer-target interactions depend on the nature of the target and on the nucleotide sequence and 3D structure of the aptamer. Aptamers occur in a wide variety of structural shapes, such as hairpins, bulges, pseudoknots and G-quadruplexes [167]. Due to these various structural shapes, aptamers can bind their targets by hydrogen bonds, hydrophobic interactions, van der Waals interactions, aromatic stacking or, in most cases, a combination thereof [89].

Crystal structures of aptamer-target complexes provide very detailed information on the interactions; they are therefore crucial for a thorough understanding of the aptamer-target binding mode. However, obtaining crystal structures of aptamer-target complexes has proven difficult, and only a few co-crystal structures have become available over the years (Table 4.1). Besides X-ray crystallography, also other techniques, such as nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), analysis using a quartz crystal microbalance (QCM) [91], isothermal titration calorimetry (ITC), Dynamic light scattering (DLS), circular dichroism (CD) [220] and small-angle X-ray scattering (SAXS) [226, 247], have been used to study aptamer-target binding. Knowledge on the target-binding mode of newly described aptamers, however, is generally restricted to the affinity of binding.

In this chapter, relevant technical issues and difficulties in the process of crystallizing aptamer-protein complexes are described. In addition, we will highlight some successful examples. Besides X-ray crystallography, several alternative approaches to investigate the molecular basis of aptamer-target interactions will be briefly discussed as well.

Table 4.1: Structures of aptamer-target complexes in the PDB database.

Target	DNA/RNA	PDB entry code	Ref.
Aptamer-protein complexes			
von Willebrand Factor Domain A1	DNA	3HXO 3HXQ	[108]
Alpha-thrombin (human)	DNA	1HUT	[210]
Alpha-thrombin (human)	DNA	1HAO 1HAP	[211]
Alpha-thrombin (human)	RNA	3DD2	[158]
Alpha-thrombin (human)	DNA	3QLP	[144]
NF-κB(p50)2	RNA	1OOA	[107]
NF-κB P50-RelB	DNA	2V2T	[182]
YmaH (Hfq)	RNA	3HSB 3AHU	[263]
a human IgG	RNA	3AGV	[200]
Enterobacterio phage MS2 coat protein complex	RNA	6MSF	[40]
Enterobacterio phage MS2	RNA	5MSF 7MSF	[243]
Enterobacterio phage MS2	RNA	1U1Y	[103]
Aptamer-small molecule complexes			
Malachite green	RNA	1F1T	[7]
Vitamin B12	RNA	1ET4 1DDY	[276]
Streptomycin	RNA	1NTB 1NTA	[283]
Biotin	RNA	1F27	[199]

Structure determination

Currently, the three-dimensional structures of many proteins, nucleic acids (e.g., riboswitches) and other biological molecules have been determined. The only methods that can provide atomic resolution structures are NMR spectroscopy, electron microscopy (electron crystallography) of two-dimensional (2D) crystals, and X-ray crystallography. Each of these methods has its advantages and limitations. In particular, NMR methods for determining high-resolution structures are largely limited to relatively small molecules (<30–40 kDa), due to complexity of the data [262, 315]. By comparison, X-ray crystallography can solve structures of arbitrarily large molecules, but it requires crystals that provide suitable quality diffraction data. Screening for conditions that yield well diffracting crystals is still a trial and error method and often requires a significant amount of effort [34]. The chance of obtaining well-ordered 2D crystals for electron crystallography is higher, when compared to obtaining 3D crystals required for X-ray crystallography; however, structure determination and

data processing from 2D crystals is still labor intensive and time consuming [227].

While providing very detailed information, crystal structures of aptamer-protein complexes have only scarcely become available. Here, we would like to describe some relevant technical issues in the process of crystallizing aptamer-protein complexes and highlight those parameters that make crystallization especially challenging.

Nucleic acid parameters

Several features of DNA or RNA play important roles in the process of co-crystallization. Apart from purity, the length and the ends (blunt or sticky) are also factors that must be explored when crystallizing nucleic acid-protein complexes.

Perhaps one of the most significant variables is the length of the nucleic acid. A general rule in protein-nucleic acid crystallization is to identify a sequence of minimal length that binds tightly to the protein [93, 96]. On the one hand, oligonucleotides that are too short will destabilize the complex, as it might limit the number of potential interactions between nucleic acid and protein, but on the other hand it may improve crystal quality, as it will reduce flanking regions that could disturb crystal contacts. Consequently, determining the minimal binding sequence of an aptamer should be considered essential. Although this is not always straightforward and can be time consuming, removing nucleotides that are unnecessary for target binding will prevent them from potentially disturbing crystal contacts.

In addition to blunt-ended nucleic acids, single- or double-base overhangs (sticky ends) are commonly explored when crystallizing complexes containing DNA or RNA [93]. It is often observed in crystals that sticky ends form crystal contacts by base-paring with complementary sticky ends, forming a pseudo-continuous double helix. In order to allow the best end-to-end packing, the overhanging bases of one strand should be complementary to the overhanging bases of the opposite strand [96]. Nevertheless, this may not be easily applicable to aptamers, because sticky ends could disrupt the tertiary structure, and hence abolish binding, of the aptamer. It has been observed, however, that the 5′ extensions from two molecules of the vitamin B_{12} RNA aptamer form a crystal contact by creating a six-base-pair duplex with two stacked adenosine-adenosine pairs [276] (Figure 4.1).

Figure 4.1: Crystal contacts between two vitamin B_{12} aptamers. The six-base-pair duplex between two vitamin B12 aptamers consists of two stacked adenosine-adenosine pairs (PDB code: 1ET4) [107]. Aptamers shown in dark and light gray, vitamin B_{12} in black.

It may also be critical to have highly purified DNA or RNA to obtain well-ordered crystals of a protein-nucleic acid complex [96]. The most common methods for purifying synthesized oligonucleotides for use in crystallization experiments are anion exchange chromatography and purification from polyacrylamide gels (PAGE). Because results of purification can vary slightly, it is sometimes desirable to further treat the purified oligonucleotide, for instance by dialyzing against an appropriate buffer.

RNA synthesis is relatively expensive; if no modified nucleotides are required, it could be cheaper and perhaps more convenient to transcribe the RNA from a DNA template. When more stable RNAs are required, one could order chemically synthesized RNA aptamers, in which all or specific uridines and cytidines are 2'-fluoro-modified. In some cases, these modified nucleotides have been used during both aptamer selection and crystallization [158, 273].

Protein parameters

A general requirement for crystallization experiments is that the protein needs to be homogeneous, i.e., highly pure (>98%) and in a low polydispersity state [64]. DLS is a commonly used technique to measure the polydispersity of a protein sample. Other parameters that should be established before setting up crystallization screens are the stability of the protein in different buffers, and whether the protein is correctly folded (CD spectroscopy) and active (activity measurements).

Protein: aptamer ratio

It is good practice to try several protein:aptamer molar ratios. In crystallization experiments, protein and DNA are usually mixed at 1:1.2 to 1:1.5 molar ratios with DNA in excess [64, 96]. The principle behind this ratio is that the concentrations of protein and DNA are frequently rough estimations and DNA could possibly not saturate all binding sites on the protein. Our advice is, therefore, to use similar ratios when setting up crystallization experiments of protein-aptamer complexes.

Crystallization screens

Screens for obtaining crystals should include a wide variety of crystallization conditions. Nowadays, many commercial screens are available from companies such as Hampton Research, Emerald BioSystems, Molecular Dimensions, Jena Bioscience and Qiagen; these also include screens specific for crystallizing nucleic acids and nucleic acid complexes. The commercial screens are undoubtedly the easiest way to initiate the first crystallization trials. For more elaborate information on protein crystallization, see [34, 64, 173]. For more elaborate information on RNA and sample preparation for crystallization of RNA and RNA-protein complexes in particular, see [53, 68, 129].

In general, for nucleic acid crystallization, it is favorable to use polyethylene glycol (PEG) or 2-methyl-2,4-pentanediol (MPD) as precipitants, rather than high salt,

because high salt may disrupt charged interactions between protein and nucleic acid. It is not possible to predict which screen will result in crystal formation, because many factors influence crystallization; variables include pH, ionic strength, temperature, protein concentration, the presence of various salts, ligands or additives, the type of precipitant and the crystallization method (hanging drop, sitting drop, dialysis, etc.) [34, 173].

Examples from literature

Over the years, only a limited number of crystal structures of aptamer-protein complexes have become available (Table 4.1), but they have provided a wealth of information. Here two aptamer-protein complexes are described that have been successfully crystallized.

Thrombin-aptamer complex

Thrombin is a trypsin-like serine protease with an important role in hemostasis, by converting soluble fibrinogen into insoluble fibrin strands. Shortly after the initial development of SELEX, thrombin-binding DNA aptamers were described [18], providing the first example of DNA binding to a protein normally not involved in DNA binding. The year after thrombin binding aptamers had first been described, a crystal structure of the thrombin-aptamer complex became available. It was obtained using a reservoir solution of 25–30% (v/v) PEG 8000, 375 mM NaCl, 0.5 mM NaN_3 and 50 mM sodium phosphate at pH 7.3, which was also used for the crystallization of a thrombin-hirugen complex [210, 261]. The crystal used for data collection grew in 2 months and diffracted to about 2.9 Å, it belongs to orthorhombic space group $P2_12_12_1$. In the same year, an NMR solution structure [162, 297] became available for the thrombin-aptamer complex as

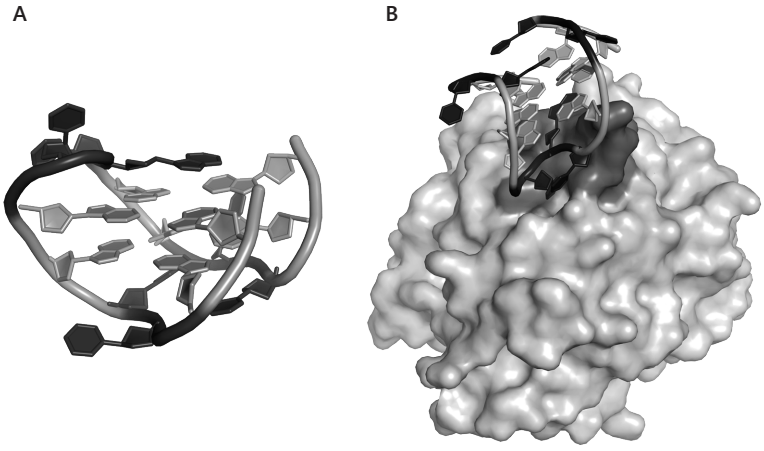


Figure 4.2: Thrombin binding aptamer. A) Double G-quadruplex shown in gray, loops in black; B) Thrombin aptamer bound to exosite 1 (dark gray) of thrombin (PDB code: 1HUT) [7].

well. The crystal structure and NMR solution structure revealed that the core of the thrombin-binding aptamers is formed by two stacked G-quadruplexes (Figure 4.2A), although the crystal and NMR solution structure show different topologies of the G-quadruplex [134]. It was also shown that the aptamer binds to exosite I of thrombin (Figure 4.2B). Recently, another crystal structure of a thrombin-aptamer complex has been published (crystallized at 20°C using 20% (w/v) PEG 20,000, 200 mM ammonium sulphate, 3% (v/v) n-propanol, 100 M sodium acetate at pH 5.8). It contains a modified thrombin binding aptamer that binds thrombin with higher affinity [144].

Anti-IgG Fc RNA aptamer

Recently, the 2.15 Å resolution crystal structure of the anti-Fc RNA aptamer in complex with the Fc fragment of a human IgG1 antibody (hFc1) was elucidated [200]. In most cases, RNA aptamers bind to their target proteins predominantly via electrostatic interactions between the negatively-charged phosphate backbones and positively-charged surface residues of proteins. To this end, the interaction between this aptamer and hFc1 is an exception, as it mainly consists of van der Waals contacts and hydrogen bonds, rather than electrostatic forces. The structure also revealed other interesting features. The RNA structure in the complex diverges greatly from its predicted secondary structure; it changes its conformation to one that structurally fits to hFc1 (Figure 4.3). The aptamer-hFc1 interaction is stabilized by a calcium ion and, therefore, binding and release of the aptamer can be controlled by either chelation or addition of calcium.

Figure 4.3: IgG binding aptamers. Two RNA aptamers bound to the Fc domain of human IgG. Residues binding the aptamer are shown in black, calcium ions (gray spheres) are indicated by arrows (PDB code: 3AGV) [315].

Crystals of the RNA aptamer in complex with hFc1 were grown by sitting-drop vapor diffusion using a reservoir solution containing 0.1 M Tris-HCl buffer (pH 8.0), 20% (w/v) PEG 1000 and 0.2 M calcium acetate. [273]. The RNA aptamer was chemically synthesized containing 2′-fluoropyrimidines and purified by PAGE. NMR analysis showed that the interaction between the Fc fragment and the aptamer has a 1:2 (Fc fragment:aptamer) stoichiometry [180]. Consequently, the aptamer was mixed with the Fc fragment in a molar ratio of 1:2.2 (Fc fragment:aptamer) for crystallization. The crystals belong to

the space group $P2_12_12$ and diffracted to 2.15 Å. Interestingly, crystal quality was improved by applying a solution stirring technique [187].

Crystallization of a streptavidin-aptamer complex

In chapter 3, we described the kinetic and stoichiometric characterization of streptavidin-binding aptamers [247]. In order to gain more insight into the molecular basis of the streptavidin-aptamer interaction, crystallization screens of the complex were conducted. Here, as an example, we will briefly describe our crystallization trials and the results obtained.

For our initial crystallization experiments, we used the commercially available crystallization screens JCSG+ Suite (Qiagen) and Crystal Screen & Crystal Screen 2 (Hampton Research, CA, USA). Streptavidin and aptamer (Figure S4.1) were mixed in a ratio of 1 streptavidin tetramer to 5 aptamers. For further experimental details, see the Supplementary text. Crystals of sufficient quality for data collection (Figure 4.4) were obtained from a high salt condition (2.0 M ammonium sulphate and 5% (v/v) 2-propanol), not containing PEG or MPD. Analysis of the 1.9 Å resolution structure showed the presence of extra electron density in the biotin-binding pocket, but this density was too small to have originated from an aptamer (Figures S4.2 and S4.3, for refinement statistics see Table S4.1).

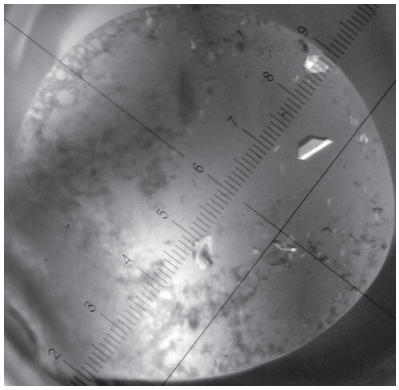


Figure 4.4: Crystal used for data collection. The crystal was obtained by vapor diffusion against a reservoir solution consisting of 2.0 M ammonium sulphate and 5% (v/v) 2-propanol (Crystal screen 2 condition 5).

In a second series of crystallization experiments, we used the Natrix & Natrix 2 screens (Hampton Research, CA, USA), which contain conditions more dedicated towards crystallization of nucleic acids and protein-nucleic acid complexes. Streptavidin and aptamer were mixed in a ratio of 1 tetramer to 2.2 aptamers on the basis of mass-spectrometry [247]. Although recently the use of chromophoric ligands was proposed to visually screen co-crystals of putative protein-nucleic acid complexes [121], we choose to conduct the crystallization screens in triplicate, using (i) streptavidin-aptamer complex, (ii) only streptavidin and (iii) only aptamer. In four conditions, crystals were obtained for both the streptavidin-aptamer complex and streptavidin alone. Crystals of the complex differed largely in size and shape from those of streptavidin only (Table S4.2), whereas no crystals grew in the same conditions containing only aptamer (pictures not shown). This indicates that the aptamer has a major effect on the crystallization process. Unfortunately, none of the crystals from the streptavidin-aptamer mix were suitable for data collection; they

were either micro-crystalline needles (1D) or plates (2D), and therefore it remains to be established whether they contain aptamer or not.

Alternative methods of investigating aptamer target interactions

Besides X-ray crystallography, alternative techniques exist to investigate the molecular basis of aptamer-target interactions. When used individually, these methods do perhaps not provide as much detailed information as a crystal structure, and they are probably not as visually appealing. However, using a combination of complementary techniques may still provide very detailed information on the interaction. Together, these techniques could be considered an alternative to X-ray crystallography, for instance, when crystallization trials fail to produce well diffracting crystals, or when no adequate equipment or expertise is available to conduct crystallization trials. Moreover, some of these techniques may provide information on the dynamics of complex formation, whereas crystallography gives a time- and position-averaged image. Since bioinformatics and experimental approaches are generally complementary, both will be discussed.

Computational approaches

The sequence of an oligonucleotide, and the intramolecular base-pairing that this sequence enables, largely defines the 3D-structure of the oligonucleotide. Base-paired regions in the 3D-structure of the oligonucleotide are thought to act as stabilizers, allowing loops and bulges to position themselves in ways suitable for target interaction [78]. The nucleotide composition of distinct aptamers that bind the same target can vary strongly, particularly in the base-paired regions; however, the overall structure could remain largely similar [14, 247, 306]. In other words, structural elements shared between oligonucleotides might therefore not be easily visible from primary sequence alignments. In addition, defining the boundaries of the minimal sequence required for target binding is not straightforward. For these reasons, identifying a recurring structure, within a number of sequenced oligonucleotides of an enriched pool, can be challenging if only sequence alignments are used.

Secondary structure predictions of clones in enriched pools may reveal recurring structural features, at any position and with deviating primary structures of oligonucleotides. These predictions could also be useful tools for identifying the minimal binding sequence. Most popular secondary structure prediction programs, including mfold [324], may be of limited value as they can only predict relatively simple hairpin structures. For the analysis of more complex aptamer structures, dedicated programs should be used, e.g., for the prediction of pseudoknots [240] and G-quadruplexes [135].

Besides secondary structure prediction, software has been developed to predict tertiary structures, varying from *ab initio* modelling to approaches requiring detailed

information on base pairing and other interactions [147]. Such software has not frequently been used in aptamer research, partly because these programs often do not provide a single prediction, but rather give numerous possibilities that each have to be scored for their relevance. For example, the MC-Fold and MC-Sym pipeline [215] that we have previously used for the generation of a 3D-model of a streptavidin binding aptamer, provided 57 widely different structural models [247]. Eventually a prediction was selected by comparing theoretical scatter of the predicted models with experimental SAXS data of the aptamer. A manual fit of the aptamer model onto a streptavidin crystal structure (Figure S3.4) showed that the aptamer could physically occupy two biotin-binding sites, and that it probably interacts with the loops that normally cover the biotin-binding pocket. This would explain the stoichiometry (2 aptamers per 1 streptavidin tetramer) and the presence of two distinct regions in the aptamer that are both essential for binding [247]. Although this modeling approach requires much more effort than standard secondary structure predictions, it can provide valuable information on the aptamer binding site and on the parts of the aptamer and the target protein that are involved in the interaction.

Experimental approaches

As mentioned above, one of the requirements for the initiation of crystallization experiments with aptamer-protein complexes is establishing the minimal binding sequence of the aptamer. As described, this can be relatively easy if a recurring sequence or structural motif is found, but binding should still be confirmed.

Alternatively, a full length aptamer may be shortened from either the 5′ or 3′ end to various lengths, by either enzymatic trimming [254] or alkaline hydrolysis [152, 304]. The resulting pool of fragments of randomly distributed lengths is incubated with the target, bound and unbound fragments are recovered as separate fractions, and run on a gel (e.g., polyacrylamide) to separate all individual fragments. From the patterns that emerge on the gel and the full-length sequence, the minimal binding sequence can be deduced. An alternative approach to determine the minimal sequence required for binding is the use of fragments of defined lengths obtained by PCR [160] or as synthetic constructs [128], and screening them for their binding capacity, e.g., by SPR. To get a better understanding of the aptamer-target interaction, nucleotides that are predicted to be important, for instance based on sequence or secondary and tertiary structure modelling, could be mutated to other nucleotides and their effect on target binding determined.

Secondary structure predictions are particularly helpful to identify the smallest DNA or RNA molecule still capable of binding the target, and may also reveal the nucleotides that are involved in target binding. On the other hand, it is also useful to obtain information on which regions or amino acids of the target are actually interacting with the aptamer. To this end, alanine scanning, replacing specific amino

acids in a protein with alanine, has been successfully used to show that two arginine residues in the protease domain of non-structural protein 3 from the Hepatitis C virus are essential for aptamer binding [113]. Although successful, this method is very laborious. Hydrogen/deuterium exchange in combination with mass spectrometry (HXMS) [313] could provide an alternative method to map sites on the protein that interact with the aptamer. Other examples of the investigation of ligand specificity are the binding of biotin to an RNA aptamer. Several small molecules, similar to biotin, or fragments thereof, were tested for their ability to bind the RNA aptamer. The results suggest that binding is not enabled by a single functionality, but rather that several interactions of different parts of biotin are required simultaneously [304]. Results of another study, in which the binding capacity of various compounds to an ethanolamine-binding aptamer was investigated, showed that the aptamer has a preference for only one functionality in the target molecule: either an ethyl- or methylamine group [225].

Other advanced analytical techniques are available for investigating aptamer-target interactions. ITC is a method used to determine thermodynamic parameters of interactions; the dissociation constant can, however, also be deduced from ITC data. SPR and QCM can be used to measure the association and dissociation rates that underlie the dissociation constant of an aptamer-target interaction. SAXS can be applied to obtain information on the size and shape of macromolecular structures; like SAXS, DLS (also referred to as Quasi-elastic light scattering) can also be used for the determination of particle sizes and shapes, but it also provides information on the diffusion coefficients in solution. CD is particularly useful in determining secondary structures of proteins and aptamer-protein complexes. Combining these dedicated analytical techniques can provide useful insights into the aptamer target interaction. A recent example for which a combination of these techniques was used (ITC, DLS, NMR and SAXS) is the transformation of a cocaine binding aptamer to one that became specific for deoxycholic acid, by only replacing a single base pair [226].

Conclusions and future outlook

Co-crystal structures of aptamer-protein complexes have provided detailed information about the interaction between the nucleic acid aptamer and the protein, and can distinguish between electrostatic interactions [144] and hydrogen-bonding interactions [200]. Crystal structures are therefore crucial for a thorough understanding of the molecular basis of the interaction. Nevertheless, obtaining crystal structures of nucleic acid-protein complexes is challenging, because a large number of parameters can be varied and numerous conditions need to be screened. In the specific case of aptamer-protein complexes, screening becomes even more challenging, because care should be taken that the 3D-structure of the aptamer remains intact during crystallization experiments. Consequently, the nucleic acid parameters (e.g., length,

presence or absence of sticky ends) are not easily varied because they could disrupt the 3D-structure of the aptamer, which should be avoided.

Despite the technical advances and the use of high-throughput crystallization methods, it will remain a major challenge to obtain aptamer-protein crystals. Improvements in NMR techniques will allow larger complexes to be analyzed at lower cost in the near future, providing an appealing alternative for crystallography. However, this requires improvements in data analysis routines, because data analysis becomes more complicated with increasing molecule size [146].

Alternative experimental approaches, such as alanine scanning, ITC, DLS, and SAXS are perhaps not as visually appealing as atomic resolution structures, but in combination with computational approaches, they may provide very useful information about the molecular basis of the interaction. In addition, some of these techniques can provide information on the dynamics of complex formation, whereas atomic resolution structures provide a time- and position-averaged image. One promising technique for obtaining high-resolution information on aptamer-protein interactions is cryo-electron microscopy (cryo-EM). Currently, cryo-EM is often used to study the structure of macromolecular assemblies, but its lower limit (~0.1 MDa) [208] is still too high for most of the aptamer-protein interactions described today. As technology advances, the lower limit might go down, and cryo-EM could become an attractive alternative for X-ray crystallography. Despite the many alternative methods to study aptamer-protein interactions, screening numerous crystallization conditions will remain, and for the time being, is the only way to obtain atomic resolution structures of aptamer-protein complexes.

Acknowledgements

This work was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research [ALW-GO-PL/08-08].

Supplementary information

Material and methods

Crystallization

Crystallization experiments were performed using the sitting-drop vapor diffusion method with the conditions provided by the JCSG+ Suite (Qiagen), the Crystal screen, Crystal screen 2, and the Natrix and Natrix 2 crystallization screens (all from Hampton Research, CA, USA).

For the crystallization experiments using the JCSG+ Suite, Crystal screen and Crystal screen 2 conditions, pure lyophilized streptavidin (M_r = 16.62 kDa; Leinco Technologies Inc., St. Louis, Mo, USA) was reconstituted in Milli-Q water to a concentration of 23.5 mg ml^{-1} (NanoDrop1000 spectrophotometer, Thermo Scientific, Wilmington, DE, USA), using a molar extinction coefficient at 280 nm of 41.940. The aptamer StrepApt5

(gggaACGCaccGATCGCAggtTtccc, Sigma; M_r = 7.96 kDa) was dissolved in Milli-Q water to a concentration of 1 mM. Streptavidin and StrepApt5 were mixed in a ratio of 1 tetramer: 5 aptamers, yielding a final protein concentration of 10 mg ml^{-1}. Drops consisting of equal volumes of reservoir solution (1 µl) and Streptavidin-StrepApt5 complex were equilibrated over 400 µl reservoirs (JCSG+) or 450 µl reservoirs (Crystal screen and Crystal screen 2) at 20°C.

For crystallization experiments using the Natrix and Natrix 2 crystallization screens, pure lyophilized streptavidin was reconstituted in Milli-Q water to a concentration of 25.5 mg ml^{-1}. StrepApt5 was dissolved in binding buffer (100 mM NaCl, 2 mM $MgCl_2$, 5 mM KCl, 1 mM $CaCl_2$ and 20 mM Tris-HCl, pH 7.6) to a concentration of 1 mM. Streptavidin and StrepApt5 were mixed in a ratio of 1 tetramer: 2.2 aptamers, yielding a final protein concentration of 10 mg ml^{-1}. The mixture was left at room temperature to allow complex formation. In addition, samples with only streptavidin or StrepApt5 were prepared by diluting them in the same way as the Streptavidin-StrepApt5 sample. Drops consisting of equal volumes of reservoir solution (2 µl) and either Streptavidin-StrepApt5 complex, Streptavidin, or StrepApt5 were equilibrated over 450 µl reservoirs (JCSG+) at 20°C.

X-ray data collection and analysis

A crystal obtained using a reservoir solution consisting of 2.0 M ammonium sulphate and 5% v/v 2-propanol (Crystal screen 2 condition 5) was used for data collection (Figure 4.3). For cryoprotection, the crystal was soaked in reservoir solution containing 25% (*w/v*) glycerol. The crystal was mounted in a cryoloop and flash frozen in a stream of nitrogen gas at 110 K. Diffraction data was collected in house at 110 K using a Microstar rotating anode (Cu) X-ray source (Bruker AXS GmbH) in combination with Helios optics (Incoatec GmbH) and a MAR345dtb detector (Marresearch GmbH). In total, 220 data images were collected with an oscillation angle of 1°. The data was processed using the iMosflm program [6] and programs from the CCP4 Suite [307]. For data collection statistics see Table S4.1.

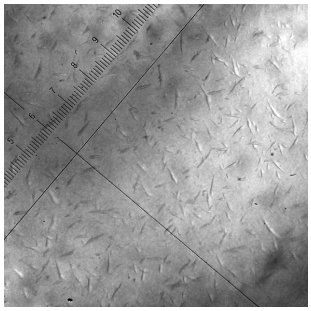
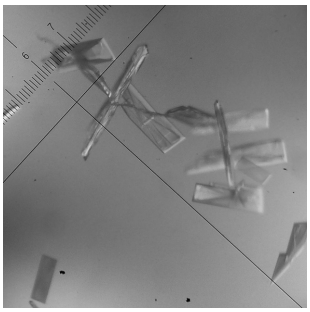
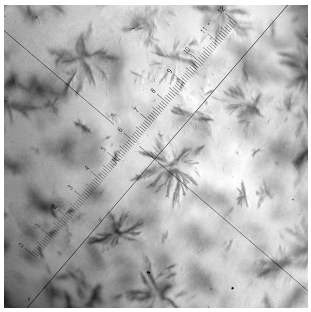
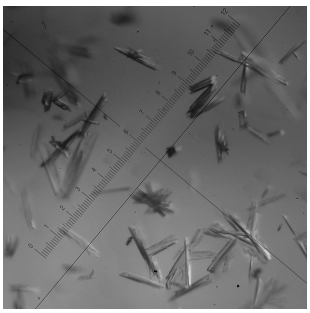
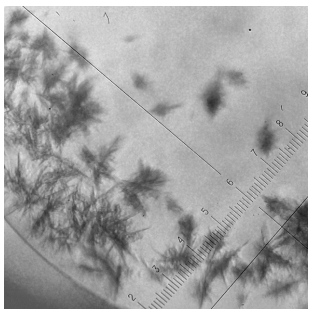
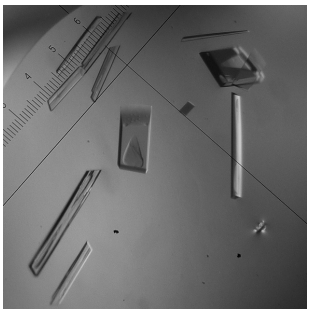
The structure of the crystallized protein was solved by molecular replacement using the program Phaser v.1.3 [171] with the native streptavidin structure (PDB code: 3RY1; [151]) as the search model from which water and solvent molecules had been removed. After rigid-body refinement of the obtained solution, the R factor was 0.396. Further refinement was done with the program REFMAC5 [188], setting aside 5% of the reflections to monitor refinement progress with the R_{free} factor, but was stopped when it was clear that no aptamer was bound to the protein. Table S4.1 lists the refinement statistics.

Table S4.1: X-ray data collection and refinement statistics.

X-ray source	In house Microstar rotating anode X-ray source (Bruker AXS GmbH)		
Wavelength (Å)	1.5418 Å		
Space group	*C*2		
Molecules/A.U	4		
Solvent content	36% (based on 64 kDa/tetramer)		
Unit-cell parameters			
a, b, c (Å)	83.9	47.6	104.7
α, β, γ (°)	90	100	90
Resolution (Å)	1.86	(1.96–1.86) *	
R_{merge}	0.101	(0.638)	
$R_{p.i.m}$	0.056	(0.338)	
Total No. of observations	138644	(18185)	
Total No. of unique reflections	32195	(4221)	
Mean *I*/σ(*I*)	7.7	(1.8)	
Completeness (%)	93.8	(85.3)	
Multiplicity	4.3	(4.3)	
Refinement			
R-factor	22.1		
R_{free}	27.1		
RMSD from target geometry			
Bond lengths (Å)	0.014		
Bond angles (°)	1.58		
Total number of atoms	3684		
Number of amino acids	475 (A: 16–136, B: 15–132, C: 15–133, D: 16–132)		
Number of water molecules	154		
Ramachandran analysis (%) [36]			
Core	88.8		
Allowed	10.3		
Generously Allowed	1.0		
Disallowed	0.0		

* Between brackets: Statistics for the highest resolution shell.

Table S4.2: Overview of crystals of streptavidin only and streptavidin-StrepApt5 complex in the same conditions.

Condition	Composition	Streptavidin only	Streptavidin-StrepApt5 complex
Natrix 1 25	0.08 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 4,000		
Natrix 1 27	0.2 M Ammonium acetate, 0.01 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 8,000		
Natrix 1 45	0.025 M Magnesium sulphate hydrate, 0.05 M TRIS hydrochloride pH 8.5, 1.8 M Ammonium sulphate		
Natrix 1 48	0.2 M Ammonium chloride, 0.01 M Calcium chloride dihydrate, 0.05 M TRIS hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4,000		

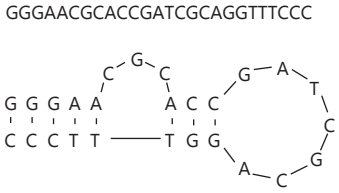


Figure S4.1: Sequence and secondary structure of StrepApt5.

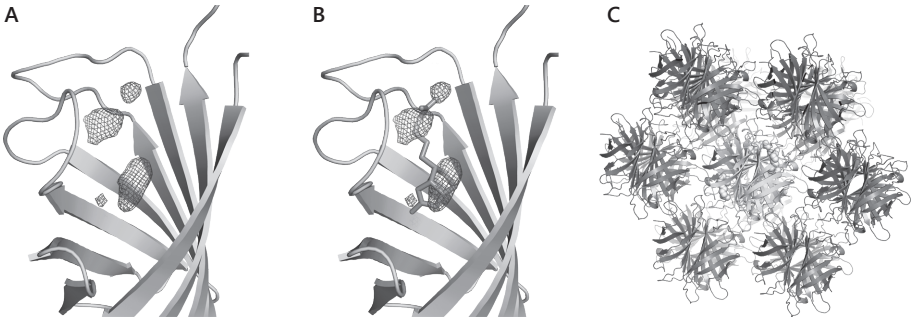


Figure S4.2: Analysis of electron density observed in the streptavidin structure obtained in this study. A) Electron density observed in the biotin-binding pocket of one of the subunits; density is present in all biotin-binding pockets; B) The observed density is too small to have originated from an aptamer, but it neither corresponds to a full biotin molecule (shown in sticks). We suspect it is a biotin analogue, used by the manufacturer to purify the protein. Unfortunately the resolution is not sufficient to identify the exact nature of this molecule; C) Symmetry-related mates of the central streptavidin show that there is no physical space for a large DNA molecule. Biotins are shown as spheres: 4 biotin molecules per streptavidin tetramer.

Figure S4.3: Streptavidin structural overlay. The structural model obtained from the collected data (unsuccessful attempt to obtain a streptavidin-aptamer structure) is shown in black. A biotin containing streptavidin structure (PDB code: 3MG5) [8] is shown in gray (biotins are hidden in this figure).

Chapter 5

Selection of a pili subunit-binding aptamer

Vincent J. B. Ruigrok, Hauke Smidt, John van der Oost, Willem M. de Vos

Abstract

Aptamers are oligonucleotide ligands that are selected for high-affinity binding to molecular targets, varying from small molecules to whole cells. However, no predictions can be made as to which target molecule will successfully yield aptamers, and no predictions can be made on which sequence will bind a pre-defined target. Therefore, selection of aptamers still remains largely a trial-and-error approach. The protein used in this study, SpaC, is a subunit of pili present on the Gram-positive bacterium *Lactobacillus rhamnosus* GG, widely marketed as a probiotic. SpaC contains a binding domain for human mucus, which we consider an advantageous feature for selection. In addition, it is an extracellular protein and presumably stable. Here we describe the overproduction and purification of the SpaC protein in *Escherichia coli*, the aptamer selection procedure and the subsequent characterization of enriched DNA sequences. A total of 49 unique aptamers were obtained in nine selection rounds; 17 of those were found to reoccur multiple times in a clone library of 85 sequences. Binding of 10 re-occurring sequences was assessed using Surface Plasmon Resonance and the results indicated that they specifically bind to SpaC.

Introduction

Aptamers are single stranded DNA (ssDNA) or RNA oligonucleotides, capable of recognizing and binding pre-defined targets with great selectivity and affinity [131, 168]; in that respect they are similar to antibodies, which are currently the most predominant affinity tool used nowadays [245]. Aptamers are selected *in vitro*, during a process named systematic evolution of ligands by exponential enrichment (SELEX [269]). A large pool of randomized DNA or RNA is incubated with a target molecule that is attached to a matrix (e.g. magnetic beads, or column material). After target exposure, any unbound oligonucleotides are washed away, bound molecules are recovered and amplified by PCR or RT-PCR; the resulting enriched pool serves as starting material for a new round of selection. Although numerous successful experiments have been described in literature [269], the selection of aptamers still remains largely a trial-and-error approach; partly because successful aptamer selection depends on properties of the target itself. Aptamers often interact via a combination of aromatic stacking, hydrogen bonding, target encapsulation, and electrostatic interactions [89]; in addition proteins are recognized via shape complementarity [89, 246]. Target molecules that enable such interactions would probably be more suitable for aptamer selection experiments, but it is not quite clear, which properties are decisive. In addition, it is not possible to predict which particular sequence will bind the target of interest. Some attempts have been made to increase the chance of successful enrichment, using partially randomized nucleic acid pools [120, 138], or pools designed to adopt a certain conformation [48, 202].

After several selection rounds the enriched pool of nucleic acid fragments is usually cloned and sequenced, in order to identify the enriched sequences. Moreover, this allows for further characterization, usually in terms of secondary structure and binding affinity. Aptamers can adopt a wide variety of structural shapes, such as hairpins, bulges, pseudoknots and G-quadruplexes [167]. Therefore, it can be challenging to accurately predict secondary structures [246]. The dissociation constant (K_D in M) of an aptamer for the interaction with its target is often measured by equilibrium methods (see e.g. in [14]). These methods cannot provide information on the rates of complex formation (k_a in M^{-1} s^{-1}) and dissociation (k_d in s^{-1}), although a lot of useful information is concealed within these parameters. Besides measuring the dissociation constant, it can also be calculated from the ratio between the dissociation and association rates ($K_D = k_d / k_a$). This also illustrates how different kinetics can result in a similar dissociation constant. To this end, surface plasmon resonance (SPR) is an advanced technique that allows measuring of the association and dissociation rates, and to subsequently calculate the dissociation constant.

Lactobacillus spp. are Gram-positive bacteria used in a wide range of food fermentations but also include groups that are marketed as probiotics [51]. One of the best known representatives of the latter category is *Lactobacillus rhamnosus* GG, which

is marketed worldwide [251]. Comparative and functional genomics studies revealed *L. rhamnosus* GG to produce pili, which are extracellular proteinaceous polymers that had previously only been reported in pathogenic bacteria [124]. The *L. rhamnosus* GG pili consist of repeating units of the 25 kDa SpaA that are decorated with 92 kDa SpaC [228, 294]. The pilin subunits are linked by intermolecular isopeptide bonds [88, 123]. Detailed analysis showed that SpaC is most often located at the pilus tip and contains a human mucus binding domain that enables *L. rhamnosus* GG to persist for longer times in the human intestinal tract. In addition, a part of the N-terminal SpaC protein sequence is similar to the type A domain of the von Willebrand factor (vWFA). The presence of such mucus-binding pili posed a new mechanism on how probiotic bacteria can interact with host tissue [124].

Although no predictions can be made on which target molecule will successfully yield aptamers, SpaC seemed a suitable candidate for an experimental approach to generate aptamers. SpaC contains a human-mucus binding domain that is already designed to bind another compound: human mucus. In addition, SpaC is an extracellular protein and stable due to the presence of intramolecular isopeptide bonds [124], which is advantageous during selection. Here we describe the production of recombinant SpaC protein, the aptamer selection procedure and the subsequent characterization of enriched sequences.

Results and Discussion

Full length SpaC protein is produced

SpaC has a theoretical mass of 91780 Da, but the protein that was produced in *E. coli* BL21(DE3) pSJS1244 cells was found to run with the same mobility as the 75 kDa marker protein (Figure 5.1). Previously, however, no such aberrant running behavior has been reported [294]. The SpaC protein was purified using the C-terminal Strep Tag II, indicating that the C-terminal region should be intact. In addition, sequence analysis

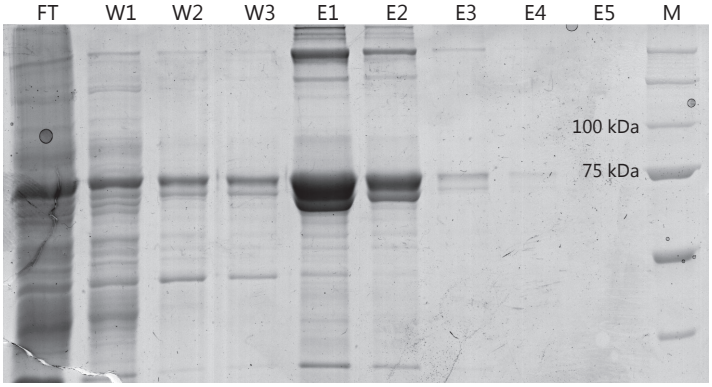


Figure 5.1: Purification of SpaC from cell free extract. Fractions are run on an 8% SDS-PAGE gel. The dark band in E1 shows purified SpaC. FT, Flow through; W, wash fractions; E, elution fractions; M, marker.

of the cloned gene revealed no abnormalities. Especially the N-terminal protein sequence (similar to vWFA) is of interest during selections; to exclude the possibility that this piece, or any other piece, of the protein is missing, the purified SpaC protein (a different batch than shown figure 5.1) was subjected to mass spectrometry analysis. This resulted in a molecular weight of 91594 Da, and although it is not an exact match, it confirms that full-length protein is indeed produced. Further analysis of the protein sequence revealed a theoretical pI of 4.85, causing the protein to be negatively charged at neutral pH; this could explain why the full-length protein runs with the same mobility as the 75 kDa marker protein. An alternative explanation might be that the presence of intramolecular isopeptide bonds in SpaC gives rise to altered formations causing aberrant migration [294]; formation of such bonds can indeed be catalyzed by the protein itself [106].

Guanine-rich oligonucleotides are enriched

Through the use of the experimental approaches described later in this chapter, we aimed to obtain ssDNA aptamers that would bind to the pili subunit protein SpaC. After nine selection rounds, 1% of the input DNA was recovered (Figure 5.2), and the enriched pool was cloned and inserts were sequenced. Assembly of 85 sequences resulted in 32 sequences that only occurred once; the 53 remaining sequences could be divided among 17 clusters, consisting of two to seven sequences (Figure 5.3), amounting to a total 49 unique sequences in the pool

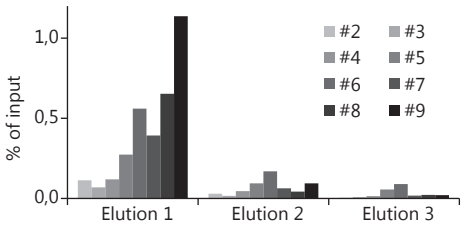


Figure 5.2: Normalized fluorescence of elution fractions of subsequent selection rounds. More than 1% of the input DNA is recovered for the first time in elution 1 of round #9.

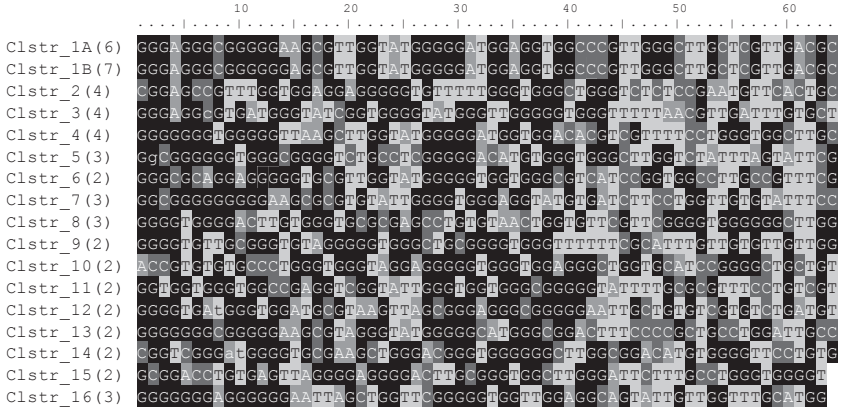


Figure 5.3: Sequences of the region between the two primer sites of the sequence clusters. The number of sequences that form a particular cluster is indicated in parentheses. Cluster 1A and 1B differ by only one nucleotide at position 14.

of 85 clones analyzed by sequencing. The difference between the most predominant sequences, cluster 1A (six sequences) and 1B (seven sequences), is only one nucleotide; 1A contains an adenine at position 14, whereas 1B contains a guanine. Although this is probably caused by a point-mutation during PCR, it could also be an exceptional example of convergent evolution.

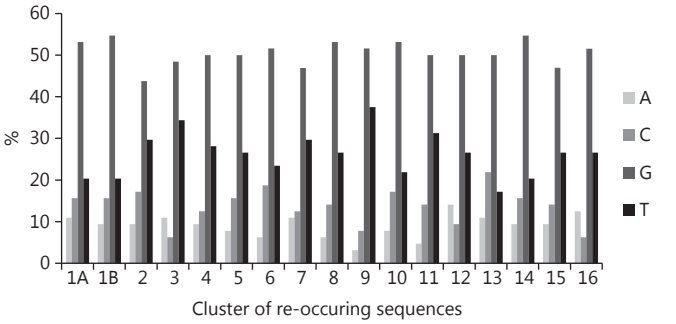


Figure 5.4: Nucleotide distribution in the sequence clusters, in the region between the two primer sites. Guanine content varies between 44–55%.

All sequences appear to have a high guanine content in the 64 nt randomized region, varying from 44–55% in the clusters (Figure 5.4). This suggests that a G-quadruplex might be formed, further strengthened by the predictions made by QGRS mapper, a tool designed among others for predicting the presence of these structures in oligonucleotides [135]. In a G-quadruplex, four guanines pair through Hoogsteen pairing, forming a planar structure [194]. Target recognition and binding could be through the loops connecting the guanines, or by electrostatic interactions of the nucleic acid backbone. The best studied example of a G-quadruplex forming aptamer is the thrombin-binding one [144]. The presence of a G-quadruplex precluded the use of secondary structure predictions, such as those made by mfold, and hence structural predictions were omitted at this stage [246, 324].

Aptamers bind specifically to SpaC

The observations that multiple sequences occur more than once in the clone library, and that they all share a guanine rich region, might hint to successful enrichment. In previous work, SELEX experiments that showed clear signs of enrichment but that failed to deliver a proper aptamer, showed no re-occurring sequences in the clone library nor did they share any particular feature (unpublished results).

No obvious sequence or secondary structure seemed to be conserved, as was the case with streptavidin binding aptamers (Chapter 3) [247]. Therefore ten oligonucleotides of selected clusters, 1A, 1B, 2–8, and 15 (excluding the primer regions) were synthesized and tested for binding in SPR experiments. Such long oligonucleotides have their drawbacks, however, once binding is confirmed efforts can be made to find out which portion of the oligonucleotide sequence is essential for binding [246]. The binding capacity was tested on two different surfaces, one containing covalently coupled SpaC and the other containing captured SpaC. Covalent coupling yields a

more stable surface, but requires harsh conditions that might damage the protein. In contrast, capture approaches work at physiological conditions but are generally less stable (Chapter 2). In addition, the Avi-tag allows for site specific biotinylation, resulting in a more homogeneous surface after capture.

It was our initial aim to first screen for binding in a qualitative way, and to subsequently characterize binding of a smaller subset of oligonucleotides in more detail. To this end all oligonucleotides were injected at a high concentration. In a second run a regeneration step was included, by injecting 1 M NaCl for 1 min. Unfortunately these SPR measurements revealed ambiguous results (Figure 5.5). There was an increase in response units, both for the covalently coupled SpaC as well as captured SpaC, which indicates binding. Furthermore, the responses to captured SpaC were much higher than to covalently coupled SpaC. The shape of the binding curve, however, provided reason to be cautious about the data. The association phase is generally not expected to follow linear behavior (although mass transport limitation could yield linear responses) and should certainly not drop below zero. During the dissociation phase complexes generally fall apart, showing a decrease in response units; this did not seem to happen in this case.

We observed some non-specific binding when the dissociation phase was followed by a buffer injection (Figure 5.6). Interestingly, the non-specific interaction appeared to be absent when the dissociation phase was followed by an injection of 1 M NaCl. A possible explanation for this observation could be that residual NaCl had not yet been fully washed away at the time a new oligonucleotide was injected. Both with and without injections of 1 M NaCl, the responses immediately dropped back to the baseline after the sample injection as is to be expected from a reference surface. In contrast, the responses did not drop back to the baseline on the SpaC coated surfaces, suggesting that the interaction is specific for SpaC.

Conclusion

In this work we have described our attempt to select aptamers specific for SpaC, a 92 kDa mucus-binding protein decorating the pili present on *Lactobacillus rhamnosus* GG. The enrichment procedure and sequence analysis indicated that aptamers were successfully enriched. In addition, the binding of several selected aptamers seemed to be specific for SpaC protein. However, we could not unambiguously demonstrate this by using SPR. Various options are open to further explore this interaction. When more SPR experiments are considered it might be useful to explore the effect of higher salt concentrations in the running buffer. Although this deviates from the original selection conditions, the regeneration with 1M NaCl provides a clue that this might be useful. Another option would be to characterize this interaction using an alternative technique, for instance by using Microscale Thermophoresis [119, 302].

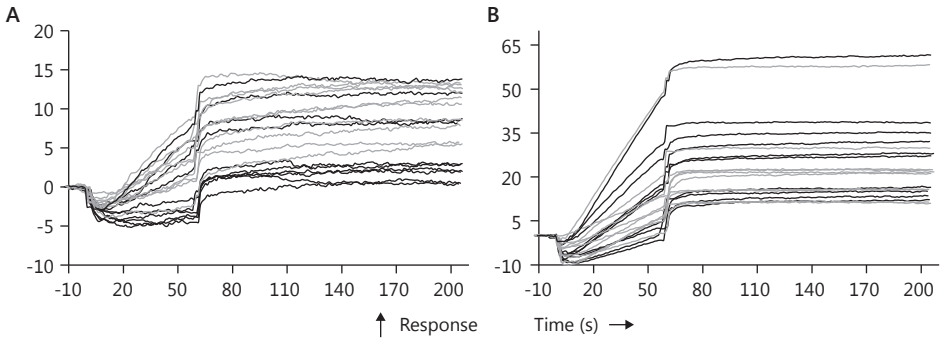


Figure 5.5: Double referenced SPR data of oligonucleotide injections (5 µM). A) Oligonucleotide binding to covalently bound SpaC (4912RU). B) Oligonucleotide binding to captured SpaC (4886 RU). Black curves were obtained without injection of 1M NaCl, gray curves were obtained with injection of 1M NaCl.

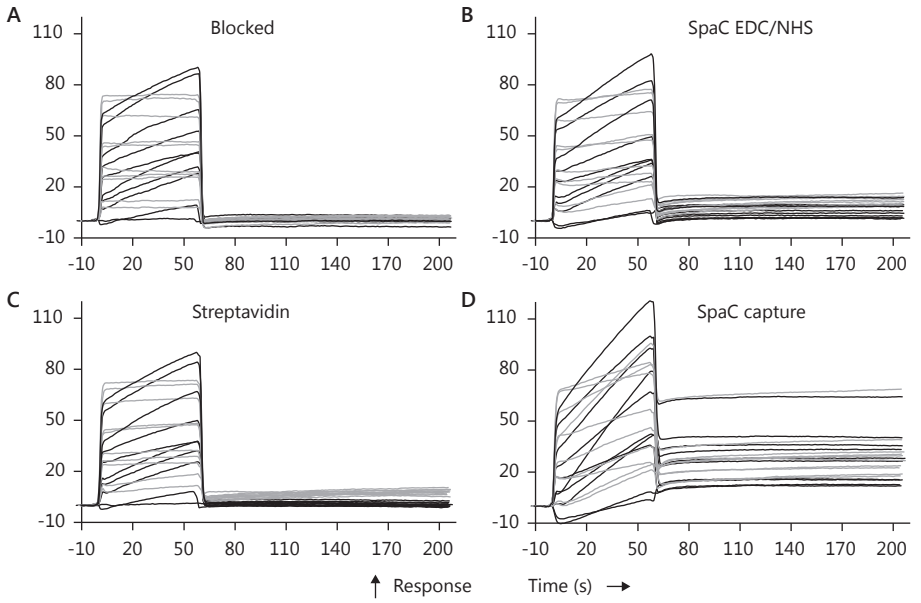


Figure 5.6: Unreferenced SPR data of oligonucleotide injections (5 µM). A) Reference surface. B) Covalently bound SpaC (4912RU). C) Streptavidin reference surface. D) Captured SpaC (4886 RU). Black curves were obtained without injection of 1M NaCl, gray curves were obtained with injection of 1M NaCl. Gray curves stay flat during sample injection over the reference surfaces (A and C), whereas black curves display non-specific binding on the reference surfaces. Both gray and black curves go back to baseline on the reference surfaces, whereas they do not on the SpaC coated surfaces (B and D), indicating that binding is specific.

Materials and methods

Cloning of *spaC* coding sequence

The coding sequence of *spaC* (LGG_00444), excluding the C-terminal cell wall sorting signal and N-terminal signal peptide, was PCR amplified from a plasmid containing a synthetic codon-optimized *spaCBA* gene cluster (F. Douillard and W.M. de Vos, unpublished data). The coding sequence was amplified in three parallel 50 µl PCRs, containing each 50 ng template DNA, 0.2 mM of each dNTP, 0.2 mM of primers BG3764 (ttcTCATGActgataacattcgcccaacc) and BG3765 (ttcGCGGCCGCcggcaaaattgcaagtgg), 1×Buffer HF and 1 Unit Phusion II (Finnzymes). The PCR program was as follows: 30 s at 98°C, 30 cycles of 10 s at 98°C, 20 s at 65°C, 45 s at 72°C, followed by 5 min at 72°C after the last cycle. Purified PCR product was subsequently digested with BspHI and NotI (sites marked in capitals) restriction enzymes, while the destination vectors, pWUR533 and pET24d-AviHisC, were digested with NcoI and NotI. Both fragments were ligated together, and *E. coli* strain XL1-blue was transformed with the resulting plasmids (pWUR533_SpaC and pET24-SpaC-AviHis) for plasmid propagation. The destination vector pWUR533 is a derivative of pET52-1b, but it contains a C-terminal HRV 3C protease site instead of a thrombin protease site, and a Strep Tag II instead of a His tag; pET24d-AviHisC contains a C-terminal Avi tag for *in vivo* biotinylation [10], and His tag for purification.

Production of recombinant SpaC

E. coli BL21(DE3) pSJS1244 was transformed with the pWUR533_Spac plasmid, encoding SpaC conjugated to StreptagII, for protein production. Fresh LB medium, containing ampicillin (final concentration: 100 µg ml^{-1}) and spectinomycin (final concentration: 50 µg ml^{-1}), was inoculated with 1% overnight culture. After three hours of incubation at 37°C, protein production was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG), to a final concentration of 1 mM. Three hours later, cells were harvested by centrifugation.

E. coli BL21-AI was transformed with pET24-SpaC-AviHis, encoding SpaC conjugated to the Avi-His tag, for protein production. This strain contains the pBirA plasmid coding for BirA biotin ligase; the vector carries chloramphenicol resistance [10, 31]. Fresh LB medium, containing kanamycin (final concentration: 50 µg ml^{-1}) and chloramphenicol (final concentration: 10 µg ml^{-1}), was inoculated with 1% overnight culture. Protein production and biotinylation was induced, three hours later, by adding isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration: 1 mM), L-arabinose (final concentration: 0.02%) and biotin (final concentration 50 µM). A 5 mM biotin stock solution was prepared in 10 mM tricin buffer pH 8.3. Three hours later, cells were harvested by centrifugation. Cell pellets were either processed immediately after harvesting, or stored at -20°C until further processing.

Purification of recombinant SpaC

Cell pellets were either resuspended in lysis buffer (150 mM NaCl and 10 mM HEPES pH 7.4) and subsequently lysed using a French press, or resuspended and lysed in B-PER (Thermo Fisher Scientific, Netherlands). Clarified cell free extract (CFE) was passed through a 0.45 µm filter before incubation with purification resin.

SpaC conjugated to Strep Tag II was purified using Strep-Tactin® Sepharose (IBA, Göttingen, Germany), which was equilibrated with lysis buffer prior to use. After incubation, the column material was washed three times with four column volumes lysis buffer. Bound protein was eluted in fractions of two column volumes using elution buffer (150 mM NaCl, 10 mM HEPES pH 7.4 and 4 mM desthiobiotin).

SpaC conjugated to the Avi-His tag was purified using HIS-Select® Nickel affinity gel (Sigma-aldrich), which was equilibrated with water and wash buffer (50 mM sodium phosphate buffer, pH 8.0, with 0.3 M sodium chloride and 5 mM imidazole). After incubation, the column material was washed three times with four column volumes of wash buffer. Bound protein was eluted in fractions of two column volumes using elution buffer (50 mM sodium phosphate buffer, pH 8.0, with 0.3 M sodium chloride and 250 mM imidazole).

SpaC bead preparation

In order to exclude the possibility of enriching aptamers binding the Strep Tag II, it was removed prior to immobilizing the protein on magnetic beads. Four units of HRV 3C protease (EMD Millipore, Billerica MA, USA) were added to 1.9 mg of SpaC protein and left overnight at 4°C. The tag and protease were subsequently removed using a Vivaspin 20 centrifugal concentrator (Sartorius, Gottingen, Germany), with a 50 kDa molecular weight cut-off. SpaC protein was coupled to M-280 Tosylactivated Dynabeads (Invitrogen) according to the procedures described in the manual. Briefly, 20 mg beads were washed with buffer B (0.1M Na-phosphate buffer pH 7.4), then 0.8 mg SpaC in 450 µl lysis buffer and 300 µl buffer C (3M ammonium sulphate in buffer B) were added to the beads. After 24hr incubation at RT, under constant upside-down rotation, supernatant was removed and beads were blocked using 1 ml buffer E (PBS pH 7.4 with 0.5% (w/v) BSA). After 1 hr beads were washed two times using buffer F (PBS pH 7.4 with 0.1% (w/v) BSA), and stored at 4°C in buffer F. Successful coupling was assessed by measuring the protein concentration before and after coupling using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

Oligonucleotides for SELEX

Unmodified primers AP10 (ataccagcttattcaatt) and AP30 (ctaactgattacgattgt) [247], primer AP60, equal to AP10 but containing 5'-Fluorescein, were obtained from Biolegio (Nijmegen, the Netherlands). Primer TER-AP30, equal to AP30 but containing 5'-poly-dA_{20}Hexaethylenglycol, was ordered from IBA (Göttingen, Germany). The

random pool (ataccagcttattcaatt-n_{64}-acaatcgtaatcagttag), purified using a reverse-phase cartridge, was obtained from Eurogentec (Seraing, Belgium).

SELEX procedure (FluMag-SELEX)

The SELEX procedure was performed essentially as described in chapter 3 [247]. Briefly, before each selection round ssDNA was heated to 95°C for 8 min, cooled on ice for 10 min and left at RT for at least 10 min. In the first round 26 pmol of the random pool was added to SpaC coated beads, and in subsequent rounds 250 µl of the ssDNA preparation [247] was added. Approximately 1×10^8 SpaC beads were washed with selection buffer (100 mM NaCl, 2 mM $MgCl_2$, 5 mM KCl, 1 mM $CaCl_2$ and 20 mM Tris-HCl, pH7.6), and after addition of the ssDNA to the beads they were incubated at 25°C with mild shaking for 30 min. Unbound oligonucleotides were removed by washing three times with selection buffer. Subsequently, bound oligonucleotides were eluted by adding selection buffer and incubation at 80°C for 5 min with mild shaking. In SELEX round 8 the selection buffer was supplemented with 0.05% BSA. In SELEX round 9 approximately 1.4×10^8 beads were used, and the selection buffer was supplemented with 0.1% BSA.

Eluted oligonucleotides were PCR-amplified in 15 parallel PCR reactions, each containing, 25 µl template (elution fraction), 1.9 mM $MgCl_2$, 0.2 mM dNTPs each, 1 µM of each primer (AP60 and TER-AP30), 1×Buffer B2, and 4 Units HOT FIREPol (Solis BioDyne) in a total volume of 100 µl. The PCR program was as follows: 15 min at 95°C, 30 cycles of 30 s at 95°C, 30 s at 52°C, 1 min at 72°C, followed by 7 min at 72°C after the last cycle. Electrophoresis on a 12% polyacrylamide (PAA) gel was used to confirm successful amplification of a DNA fragment of the correct size.

From the second selection round onwards fluorescence in the starting sample, the non-bound DNA fraction, the wash fractions, and the eluted DNA was measured in 96-well microtiter plates (Greiner bio-one) using a SpectraMax M2 (Molecular Devices).

Cloning and sequencing of enriched fractions

DNA eluted in SELEX round 9 was amplified, according to conditions described above, using unmodified primers AP10 and AP30 and ligated into pGEM®-T (Promega). *Escherichia coli* XL1-Blue cells (Stratagene) were transformed with the vector constructs, subsequently 96 colonies were randomly picked and vector inserts were sequenced using the T7 forward primer (GATC, Konstanz, Germany).

Surface Plasmon Resonance

SPR experiments were performed on a Biacore 3000 system (BIACORE, Uppsala, Sweden), at a constant temperature of 25°C. Sensor chips (CM5), 0.1 M *N*-hydroxysuccinimide (NHS), 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)

carbodiimide hydrochloride (EDC), 1 M ethanolamine hydrochloride (pH 8.5) and 10 mM sodium acetate pH 4.0, were from GE Healthcare (Uppsala, Sweden). Streptavidin was obtained from Leinco Technologies Inc (St. Louis, MO, USA). A flow of 5 µl min^{-1} was used during all coupling and capture steps.

On one CM5 chip, biotinylated SpaC was covalently bound in fc 2, to a density of 4912 Response Units (RU), and captured on a streptavidin surface in fc 4, to a density of 4886 RU. The surface in fc 2 was activated by injecting 35 µl of a 1:1 mixture (v/v) of EDC and NHS. Then 100 µl SpaC, approximately 50 µg ml^{-1} in 10 mM sodium acetate pH 4.0, was injected. Finally, the surface was blocked by an injection of 35 µl 1 M ethanolamine hydrochloride. A reference surface was created in fc 1 by activating and blocking the surface. A streptavidin surface was created in fc 3 and 4 using the same approach, by injecting 150 µl streptavidin, diluted to approximately 50 µg ml^{-1} in 10 mM sodium acetate pH 4.0, after activation of the surface. Subsequently, SpaC was captured in fc 4, and fc 3 was left empty, to serve as reference surface.

The 64 nucleotide region (in between the primers) of ten of the enriched oligonucleotides was synthesized (Biolegio, Nijmegen) and tested for binding, using selection buffer as running buffer. To this end, 30 µl of a 5 µM solution of each synthetic oligonucleotide was injected (using the kinject command) at a flow of 30 µl min^{-1}, followed by a dissociation phase of three min. In one set of experiments the dissociation phase was followed by a 15 µl buffer injection, in another set of experiments it was followed by a 15 µl injection of 1 M sodium chloride. Data was processed and analyzed using Scrubber (BioLogic Software, Campbell, Australia).

Acknowledgements

This work was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research [ALW-GO-PL/08-08] to HS, and the ERC grant Microbes Inside (grant 250172), to WMV. The authors wish to thank Esther van Duijn for Mass spectrometry analysis of SpaC protein.

Chapter 6

Assessing the dynamics of aptamer selection by ultra-deep sequencing

Vincent J.B. Ruigrok, Javier Ramiro-Garcia, Carel Fijen, John van der Oost, Hauke Smidt

Abstract

Despite the successes in aptamer selection towards specific and sensitive binding molecules, the dynamics of the *in vitro* selection process are not fully understood. Experimental conditions could be further optimized, instead of being based on practical considerations and convenience, as it is often the case. In addition, not necessarily the tightest binder is enriched but rather those with the most rapid association rate. Here we describe a multiplexed high throughput sequencing approach, in which 70 samples were mixed, derived from eight distinct aptamer selection experiments, in order to get a more thorough understanding of the dynamics of aptamer selection. We also compared the effect of various elution methods. To our knowledge, this is the first time that so many samples of multiple selection experiments are combined in one study. Over 84 million paired-end reads were obtained, and a preliminary analysis of the data shows a decrease in α-diversity for successful aptamer selection experiments. Furthermore, the comparison between thermal and affinity elution methods (β-diversity) clearly shows that different oligonucleotides are enriched. A small-scale comparison of two clone libraries further suggests that the affinity elution method specifically enriches rapid binders.

Introduction

Aptamers are single stranded DNA or RNA oligonucleotides selected to bind a predefined target. Since their initial development two decades ago [58, 285] the field of aptamer research has matured, and aptamers have become available for numerous targets.

Systematic evolution of ligands by exponential enrichment (SELEX) is the procedure by which aptamers are selected from a large pool (10^{14}–10^{15} variants) of randomized nucleic acids. Over the years, various modified SELEX procedures have been described and successfully used [269]. Generally a SELEX procedure consists of multiple subsequent selection rounds, and starts with exposing RNA or ssDNA oligonucleotides to a target molecule that has been immobilized on a surface matrix, e.g. column material [305] or magnetic beads [268]. Non-binding oligonucleotide molecules are washed away, after which bound molecules are recovered by either specific (affinity elution) or non-specific (e.g. heat elution) elution methods. Recovered oligonucleotides are amplified by PCR or RT-PCR and made single-stranded again in order to be used as starting material in a next round of selection. After successful enrichment is observed, the enriched pool is cloned and sequenced. Often, binding parameters of several individual sequences are assessed to identify the tightest binder.

Despite the successes in aptamer selection, the dynamics of the selection process are not fully understood and hence experimental conditions could perhaps be further optimized, to aid more rapid selection procedures and to yield tighter binders. Nowadays experimental conditions are often based on practical considerations, such as available expertise and experiences in the past. Convenience, such as the availability of a target-coated matrix plays a role as well; column materials with a large variety of targets are commercially available, whereas target coated magnetic beads are not widely available. Conditions like the amount of beads or column material, the amount of starting nucleic acid, the elution method and the number of selection rounds used during a SELEX experiment also influence the final outcome of a SELEX experiment. In addition, not necessarily the tightest binder is enriched but rather the binder with the most rapid association rate (chapter 3) [247].

In recent years, unprecedented advances in sequencing technologies [178] have led to a variety of new sequencing applications [255]. As might be expected, the SELEX procedure has been subjected to high throughput sequencing approaches as well. Such studies, however, primarily aimed to develop more rapid selection methods [38, 102], as opposed to studying the actual selection process. One study, also aiming for an improved selection method, provided somewhat more insight in the relative abundance levels of dominant sequences throughout several selection rounds [253]. However, more informative methods are available for the analysis of the diversity and structure of e.g. microbial communities [159]. Generally, diversity within a community is characterized by α-diversity, which gives an estimate of the number of species

present (species richness). Furthermore, relatedness between communities can be characterized using β-diversity parameters, which are often based on the number of species shared between communities [222].

In the present study, we describe a multiplexed high throughput sequencing approach, aiming to study the diversity and composition of enriched oligonucleotide pools, using α and β-diversity estimates originally developed for ecological studies. In total, 70 samples of subsequent selection rounds of eight distinct SELEX experiments were pooled and sequenced. In four of these SELEX experiments, aptamers were selected for streptavidin and hemoglobin, using affinity elution and thermal elution. Samples of the four remaining SELEX experiments are not discussed here for IP-related reasons. Studying these samples provides more insights in the dynamics of aptamer selection, and allows us to compare the effect of different elution methods on the outcome of selection targeting a similar target.

Results and discussion

Aptamer selection for streptavidin and hemoglobin

In total, six selection rounds were performed for streptavidin using affinity elution. For hemoglobin, six rounds were completed for the thermal elution selection, while four selection rounds were completed using affinity elution. For streptavidin, enrichment was first observed after four selection rounds and continued to increase in rounds five and six, when more than 30% of the input DNA was recovered (in Elution 1) based on normalized fluorescence of DNA in the elution fractions (Figure 6.1A). PCR reactions on DNA eluted in round four and five, however, yielded only very small amounts of DNA, resulting in decreased starting amounts of material in round five and six (data not shown).

During the selection of hemoglobin aptamers, using non-specific thermal elution, enrichment was observed in round three, although it should be noted that,

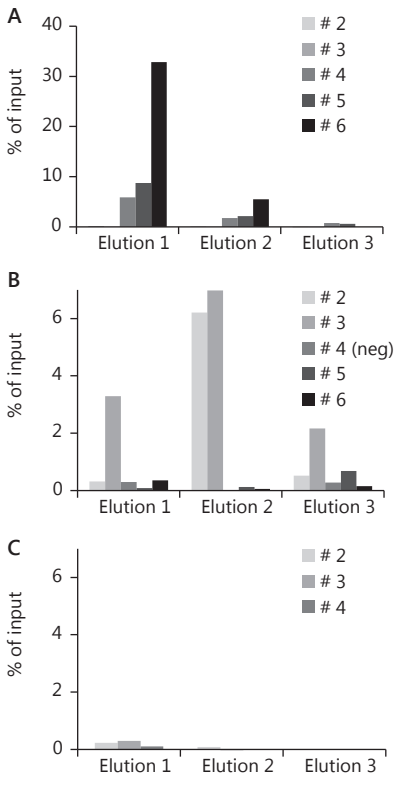


Figure 6.1: Normalized fluorescence of elution fractions in subsequent selection rounds. A) Enrichment detected from round four onwards during streptavidin affinity elution, B) no recovery after negative selection (neg) during the hemoglobin thermal elution selections, C) During hemoglobin affinity elution there is no enrichment at all. Scales for both hemoglobin experiments are the same, for easy comparison.

unexpectedly, the highest recovery was found for elution 2 (Figure 6.1B). After a round of negative selection, by exposure to 'empty' streptavidin beads, however, no DNA was recovered anymore after round 3. In addition, no enrichment was observed for the hemoglobin selection when affinity elution was used (Figure 6.1C). This suggests that streptavidin binders were being enriched, instead of hemoglobin binders, when using thermal elution. Although both hemoglobin SELEX experiments did not result in an aptamer being enriched we have included the selection samples, to possibly identify aptamer sequences enriched throughout subsequent selection rounds. Enrichment of specific sequences can be accurately traced back in large sequence datasets [102].

Diversity decreases in subsequent selection rounds

Through the use of a multiplexed high throughput sequencing approach, we aimed to gain a more fundamental insight in the dynamics of the SELEX procedure. A total of 70 samples, obtained during eight SELEX experiments, were mixed in one equimolar mixture and sequenced on a HiSeq2000. In total 84.511.225 pairs of reads were obtained. The overall quality of the sequence data was good and the complete random region was covered. Theoretically, each sample is expected to be represented by ~1.2×10^6 reads. From the actual read distribution (Table 6.1), however, it is apparent that samples with a pool length of 100 nt (samples number 1–27 and 63) contain more reads, averaging $1.8\pm0.7\times10^6$, than those with a 76 nt pool length (samples number 28–62 and 63–70), averaging $4.3\pm1.4\times10^5$. However, this poses no problem for subsequent analysis, because the reads are distributed quite evenly within samples of each SELEX experiment.

Streptavidin-binding oligonucleotides were enriched in two different ways. In one SELEX experiment, described in chapter 3, binding oligonucleotides were recovered in a non-specific manner, by thermal elution. In this study, binding oligonucleotides were recovered in a specific manner, by affinity elution [247]. Rarefaction analysis of selection samples of both streptavidin SELEX experiments revealed indeed that the α-diversity decreased as selection progresses, indicated by the lower asymptote of the rarefaction curves of later selection rounds (Figure 6.2 A and B). As a consequence of using a database for OTU picking, containing only those sequences represented by a minimum of 100 reads across the entire dataset, only a small number of the reads in the earlier selection rounds are incorporated in the analysis. This resulted in a limited number of points in the rarefaction curves, but nevertheless it is clear that these curves are heading towards a more diverse composition. Interestingly, there is a clear difference between both elution methods. Selection using thermal elution resulted in a more gradual decrease in α-diversity, whereas selection using affinity elution is characterized by a sudden decrease in α-diversity after round 4 (S3 to S4 in Figure 6.2B). Moreover, the affinity elution samples appeared to be less diverse than those obtained by thermal elution, although the latter might not yet have stabilized.

Rarefaction curves of both hemoglobin SELEX experiments (Figure 6.2 C and D) indicate that α-diversity does not significantly decrease in subsequent selection rounds; this observation was especially valid for thermal elution. Although the α-diversity was somewhat decreasing for the affinity elution experiment, only a small number of reads (max 5%) were re-occurring often enough to be included in the data analysis. The small number of reads included in the analysis and the high diversity indicated by rarefaction, which is likely to be an underestimation, might explain why these selection experiments did not show enrichment using fluorescence measurements. This does not exclude the possibility that aptamers are enriched for

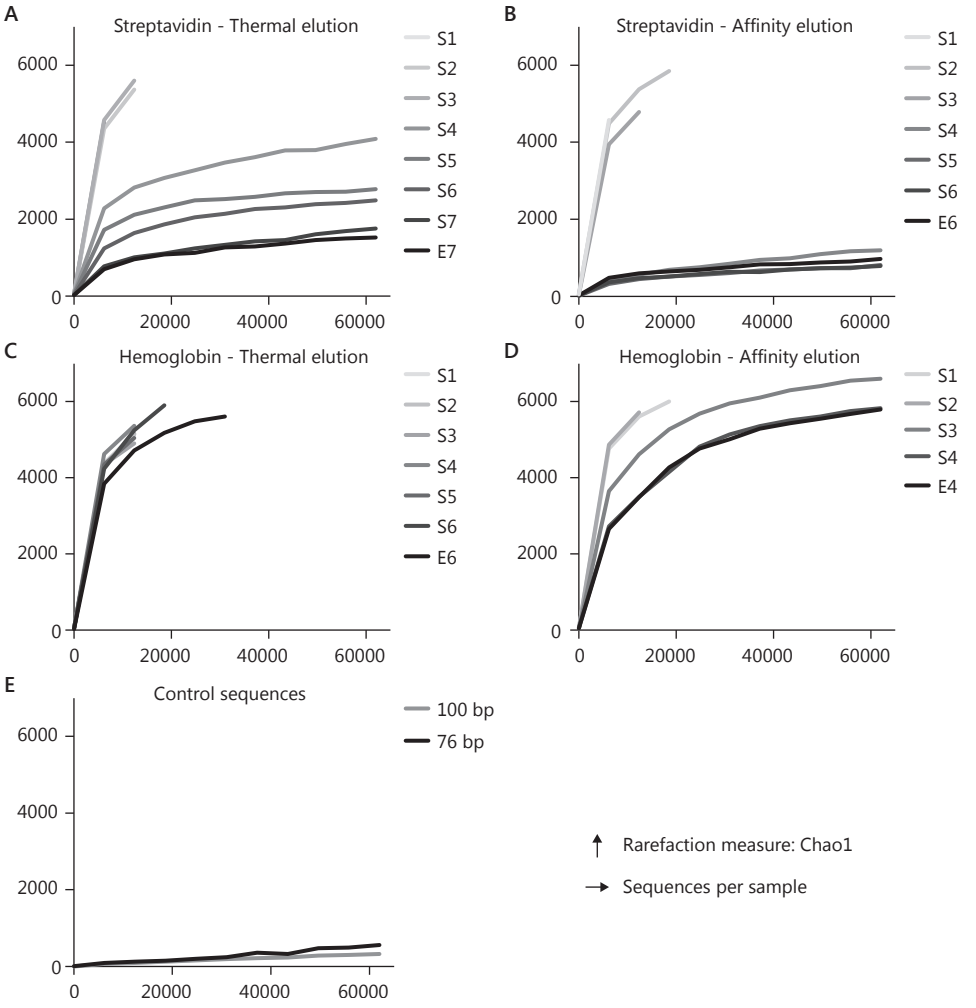


Figure 6.2: Rarefaction analysis of aptamer selection samples. A) Diversity decreases in subsequent selection rounds B) At S4 diversity is stabilized. C and D) No obvious decrease in diversity is observed. E) Control sequences for both pool lengths. S indicates starting material, and was either the synthesized random pool (S1), or eluted material from the preceding selection round that was PCR amplified and made single stranded (S2–S7). E indicates the final elution.

hemoglobin, however, more detailed analyses are required to address this [102]. The two samples that were included as controls behaved as expected and showed a low diversity (Figure 6.2E), however, approximately 10–25% of the control reads were not included in the analysis. This suggests that either the template was not completely homogeneous or, more likely, that the error rate during sequencing is high.

Rapid binders are enriched by affinity elution.

A more detailed comparison (Pearson Principle Coordinate Analysis) of both streptavidin selection experiments revealed that communities of the early selection rounds of both selection methods are more similar to each other than the enriched pools of later selection rounds (Figure 6.3). Forward and reverse sequencing reads were kept separate during this analysis. The forward reads for the samples of the first two rounds of selection (marked by S1* and S2*) were slightly dissimilar from the other early selection rounds (S1–S3), however, this could be due to the small number of reads that were included in the analysis. Furthermore, samples of later selection rounds of thermal and affinity elution samples, formed separate clusters respectively, indicating that different oligonucleotides were enriched by both elution methods.

In order to gain a more detailed understanding of the differences caused by the different elution methods a small-scale comparison was initiated. To this end, the enriched pool of streptavidin binding oligonucleotides, recovered using affinity elution, was cloned after round four (Sample S5). Signs of enrichment were detected for the first time at the end of round four by fluorescence measurements, even though it should be noted that rarefaction analysis of Illumina sequences showed already a large drop in α-diversity at the start of round four (S4; Figure 6.2B). Moreover, the starting material of round four also clusters with the enriched fractions of later selection rounds, indeed suggesting that enrichment had already commenced.

In total, 96 clones were sequenced, of which 84 were used for subsequent analysis. Assembly of individual sequences yielded eight clusters that each contained two to 26 fully identical sequences, whereas six sequences occurred only once, resulting in a total of 14 unique sequences. Structural predictions of all unique sequences (clusters and singlets) were made using mfold [324], and in all sequences a similar structure of a stem-bulge-stem-loop was predicted, in line with previous findings [247].

The sequences predicted to form the stem-bulge-stem-loop structures were manually aligned according to their positions within the predicted structures (Figure 6.4). Despite the overall structural similarity some small variations could be observed: loop sizes vary between five and six nucleotides, and the bulges contain either three or five nucleotides. In particular, the nucleotides in the bulges (CGC) and loops (CGCA) are well conserved, whereas the actual compositions of the stems do not seem to be important as long as proper Watson-Crick base pairing is maintained. From the alignment and the small differences between sequences it was possible to

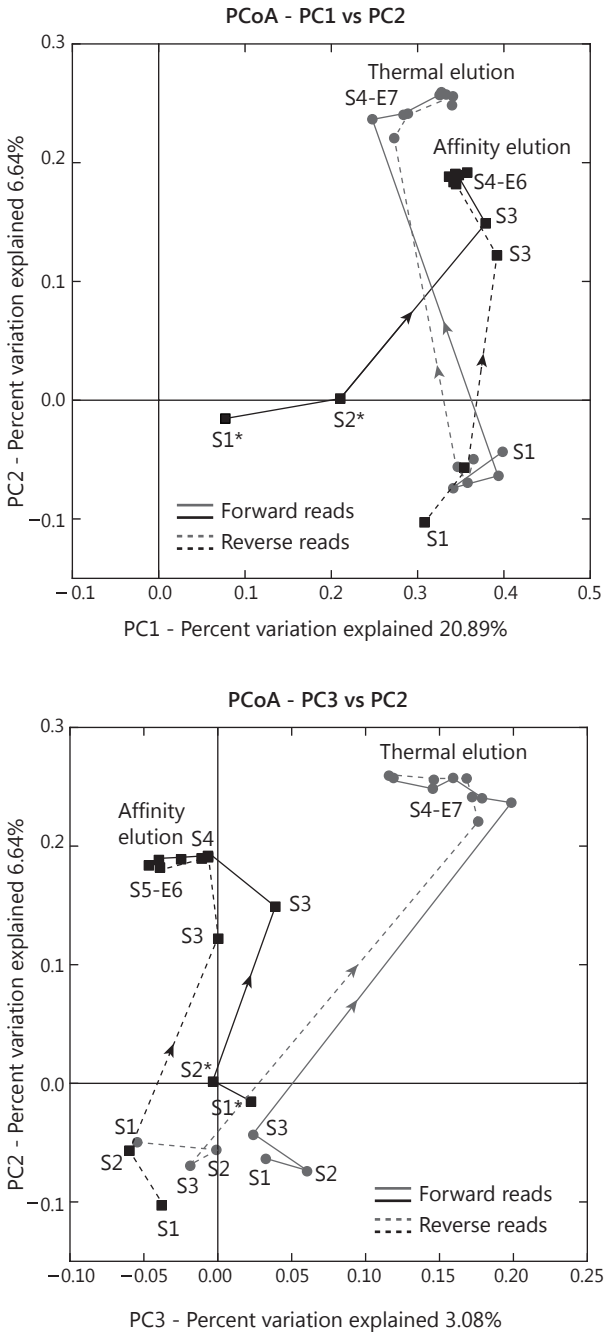


Figure 6.3: Pearson Principle Coordinate Analysis of Streptavidin selection rounds. Results of both forward reads (continuous lines) and reverse reads (dashed line) are shown, thermal elution samples are indicated as gray circles and affinity elution samples as black squares. Arrowheads indicate the direction of samples of subsequent selection rounds, toward the thermal elution cluster and affinity elution cluster, respectively. * Indicates forward reads that are slightly dissimilar.

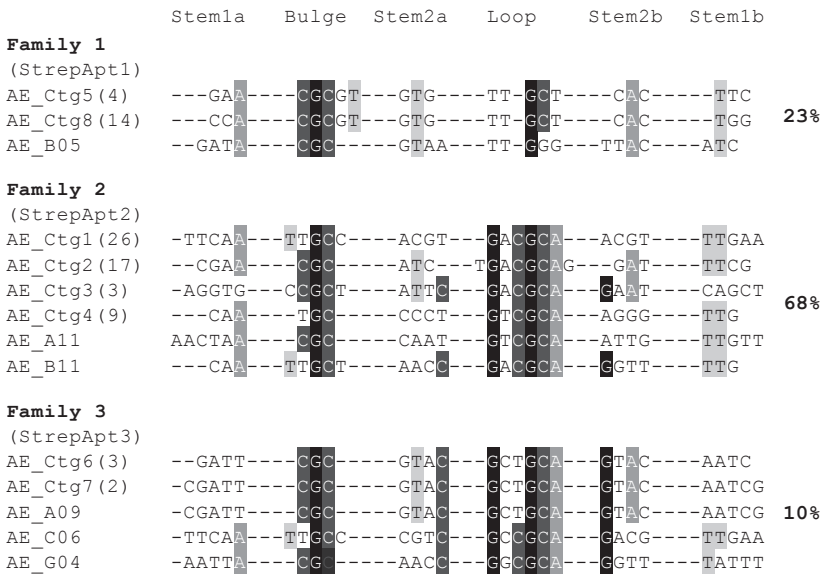


Figure 6.4: Alignment of stem-bulge-stem-loop sequences enriched by affinity elution. AE stands for affinity elution, Ctg is a cluster of identical sequences. The number of sequences in a particular cluster is indicated in parenthesis. Percentages indicate the occurrence of a certain family within the clone library.

distinguish three aptamer families that are similar to three of the five families that we described in chapter 3 [247]. Some sequences are even identical to those we found before, for instance AE_Ctg2 of this study.

After four rounds of affinity elution, the number of unique sequences has dropped to 14, whereas thermal elution resulted in 31, after seven rounds of selection. Moreover, in the clone library of the thermal elution SELEX, five distinct families of streptavidin-binding aptamers were identified (Figures 3.2 and 3.3). Each family has their own distinct kinetic behavior, and it was noted that the three aptamer families with a rapid rate of complex formation dominated the clone library. In the clone library of round four of the affinity elution SELEX, however, we could identify only 3 families, similar to those we described before. Interestingly, these three families all correspond to the three families that dominated the thermal elution SELEX. Hence, while affinity elution is expected to result in aptamers more specific to a pre-defined target, it also seems to specifically enrich for rapid binders. This is not completely unexpected, because during affinity elution two distinct binding stages can be distinguished. First, binding of oligonucleotides to the target coated matrix. Second, re-binding of bound oligonucleotides to uncoupled target present in the affinity elution buffer. It seemed likely that during both of these events rapid binders have an advantage, but this has not been so clearly demonstrated before.

Conclusion

The multiplexed high throughput sequencing approach presented in this chapter has shown to be helpful in obtaining diversity indices of a large number of selection samples from eight distinct SELEX experiments. As is to be expected, diversity dropped during the course of selection, clearly demonstrated by the rarefaction analysis of the streptavidin SELEX experiments. In addition, community structure analysis of two elution methods, used for enriching streptavidin-binding oligonucleotides, showed that different oligonucleotides were enriched. Furthermore, the small scale comparison of two clone libraries, originating from these two distinct enrichment experiments, provided a clear indication that affinity elution methods specifically enrich for rapid binders. As a next step, representation of aptamer families throughout the selection process in the HiSeq2000 dataset should be explored in more detail, especially because we can relate a certain family type to kinetic behaviour for the streptavidin-binding aptamers. Possibly, newly developed high-throughput methods to identify sequence structures [94] could help us to zoom in on the point at which the slow binders take the defeat of the rapid binders, a turning point for the kind of properties that are beneficial for enrichment.

Materials and methods

Hemoglobin bead preparation

Hemoglobin (Sigma, CAS: 9008-02-0) was biotinylated according to procedures and reagents provided in the EZ-Link Sulfo-NHS-LC-Biotinylation Kits (Thermo Fisher Scientific, Etten-Leur, The Netherlands). To 1.2 mg hemoglobin (in 2 ml PBS), 37 µl of a 10 mM Sulfo-NHS-LC-Biotin was added and left rotating at RT for 1 hour. Uncoupled biotin was removed by using a Vivaspin 20 centrifugal filter with a 30 kDa cut-off (Sartorius), in three consecutive rounds of centrifugation (10–15 min, 4500 rpm at 4°C) and buffer addition. Coupling of biotin to hemoglobin was assessed using the HABA assay, provided with the biotinylation kit. Biotinylated hemoglobin was captured on streptavidin magnetic beads (Dynabeads® M-280, Invitrogen). Prior to capture, beads were washed three times with PBS. Beads were left to rotate at RT for 1 hour after addition of biotinylated hemoglobin.

SELEX samples

Samples from streptavidin thermal elution were derived from another SELEX experiment, described in chapter 3 [247]. Hemoglobin SELEX samples, both thermal and affinity elution, as well as streptavidin affinity elution samples were obtained for this work by methods exactly as described previously [247]. For the affinity elution method, however, elution of bound oligonucleotides was done by using selection buffer containing 0.6 mg ml^{-1} hemoglobin or 0.6 mg ml^{-1} streptavidin, and subsequent incubation at room temperature for 15 min. A list of all samples is found in Table 6.1.

Samples 28–70, except for 63, were obtained from collaborators (Department of Environmental Biotechnology, Centre for Environmental Research Leipzig-Halle GmbH), and will not be discussed further in this chapter.

Cloning, sequencing and structure analysis

DNA eluted in round four of streptavidin affinity elution selection was amplified, using unmodified primers AP10 and AP30, and ligated into pGEM®-T (Promega). *Escherichia coli* XL1-Blue cells (Stratagene) were transformed with the vector constructs, subsequently 96 colonies were randomly picked, and vector inserts were sequenced using the T7 forward primer (GATC Biotech, Konstanz, Germany). The DNA folding form of mfold [324] was used for secondary structure prediction with use of the default settings except for temperature (25°C) and ionic conditions: Na^+ (100 mM) and Mg^{2+} (2 mM).

Barcodes for high throughput sequencing

Eight nucleotide error-correcting golay barcodes were selected in such a way that they were at a mutation distance 3. Of the resulting 727 barcodes, 70 barcodes were selected in such a way that the distribution of each nucleotide at each position is as close to 25% as possible in order to meet technical requirements of the Illumina HiSeq2000 protocols. Each barcode was assigned to one of 70 samples (see Table 6.1). The resulting 140 primers were obtained from Biolegio (Nijmegen).

Equimolar mixture for high throughput sequencing

Samples 28–70 (except 63) were PCR amplified using unmodified AP10 and AP30 or AP20 (65–70, agattgcacttactatct) primers in two PCR reactions of each 100 µl, containing 1 µl template, 1.9 mM $MgCl_2$, 0.2 mM dNTPs each, 1 µM of each primer, 1×Buffer B2, and four units HOT FIREPol (Solis BioDyne). The PCR program was as follows: 15 min at 95°C, 30 cycles of 30 s at 95°C, 30 s at 52°C, 1 min at 72°C, followed by 7 min at 72°C after the last cycle. Both reactions were pooled and subsequently purified using the GeneJET™ PCR Purification Kit (Fermentas), using the optional steps for molecules <500bp.

Barcodes were added by PCR in three PCR reactions of each 100 µl, containing approximately 10 ng template, 1.9 mM $MgCl_2$, 0.2 mM dNTPs each, 1 µM of each primer, 1×Buffer B2, and 4 Units HOT FIREPol (Solis BioDyne). The PCR program was as follows: 15 min at 95°C, 5 cycles of 30 s at 95°C, 30 s at 52°C, 1 min at 72°C, followed by 25 cycles of 30 s at 95°C, 30 s at 60°C, 1 min at 72°C followed by 7 min at 72°C after the last cycle. All three reactions were pooled and purified using the GeneJET™ PCR Purification Kit (Fermentas), using the optional steps for molecules <500bp. DNA was eluted in 50 µl MQ-water.

DNA concentrations were measured on a Qubit fluorimeter (Invitrogen, Bleiswijk,

The Netherlands) using the Qubit® dsDNA BR Assay Kit (Invitrogen). Subsequently, equimolar amounts were pooled. To avoid pipetting of very small volumes, the calculations were done in such a way that 0.75 µl of the highest concentrated sample was added to the equimolar mixture.

Multiplexed high throughput sequencing and data processing

The resulting equimolar mixture was concentrated in a speedvac prior to shipping to GATC Biotech (Konstanz, Germany) for sequencing. Tagged adapter sequences were added at GATC Biotech, in order to prepare the equimolar mix for paired end sequencing on a HiSeq 2000 (one lane). Sequences were sorted by tag, omitting those that contain one or more mismatches in the tag region from further analysis. Distribution of the reads is shown in Table 6.1. Tag-sorted data was further processed and analyzed using QIIME [28]. Because *de novo* picking of Operational Taxonomic Units (OTU) is not feasible with such a large data set, a reference database was created and used for OTU picking. In the database only sequences were included that are present more than 100 times within the whole dataset. The resulting OTU table contains 7.923 OTUs, representing a total of 49.863.130 sequence reads. This OTU table was used for subsequent diversity analysis (without assigning taxonomy or phylogenetic tree building).

Acknowledgements

This work was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research [ALW-GO-PL/08-08]. The authors wish to thank Detmer Sipkema for helpful discussions on data analysis topics.

Table 6.1: Sample information, sequence read distribution and read numbers used for analysis.

Sample					Tag sequence	Pool (nt)	Forward read	Reverse read	Used in analysis		% of reads used	
									For	Rev	For	Rev
1	Streptavidin	Thermal elution	Start rnd	1	AACCAACG	100	874.261	920.984	4999	12232	0,57	1,33
2				2	AACCAGAA	100	1.037.658	1.124.556	5931	16226	0,57	1,44
3				3	AACCATGC	100	1.035.882	1.104.454	6207	15372	0,6	1,39
4				4	AACCGCCA	100	1.266.246	1.330.986	331132	369172	26,15	27,74
5				5	AACCGGTT	100	2.307.385	2.348.734	1721563	1880887	74,61	80,08
6				6	AACCTAGA	100	2.685.035	2.718.364	2187676	2312630	81,48	85,07
7				7	AACGAAGT	100	2.743.379	2.753.127	2247819	2390344	81,94	86,82
8			Elut	7	AACGACTA	100	3.136.441	3.154.407	2520478	2712633	80,36	86,94
9	Streptavidin	Affinity elution	Start rnd	1	AAGAACCA	100	816.869	822.748	4162	10022	0,51	1,22
10				2	AAGAATGG	100	2.128.859	2.139.143	8376	19901	0,39	0,93
11				3	ACCAAGAT	100	937.413	857.049	9953	14150	1,06	1,65
12				4	ACCAGACC	100	1.788.445	1.715.084	1201850	1288774	67,2	75,14
13				5	ACGCTAAT	100	2.528.396	2.515.405	2067837	2195257	81,78	87,27
14				6	AGAACCGG	100	3.433.389	3.423.678	2633607	2892181	76,71	84,48
15			Elut	6	AGAACTCT	100	2.145.333	2.131.134	1771028	1852726	82,55	86,94
16	Hemoglobin	Thermal elution	Start rnd	1	AGAGGTCC	100	1.076.855	992.637	6745	12933	0,63	1,3
17				2	AGGTCGAA	100	1.267.505	1.178.806	6475	13403	0,51	1,14
18				3	ATTGCATT	100	1.060.307	883.472	6813	12627	0,64	1,43
19				4†	CATAATTA	100	1.944.824	1.932.058	5080	16212	0,26	0,84
20				5	CATGATGC	100	1.708.233	1.656.882	5314	16034	0,31	0,97
21				6	CCAAGAGT	100	1.625.796	1.565.088	12049	23767	0,74	1,52
22			Elut	6	CCAGAGCG	100	1.339.154	1.306.730	21491	32484	1,6	2,49
23	Hemoglobin	Affinity elution	Start rnd	1	CCAGCTGG	100	1.807.394	1.769.239	8223	19656	0,45	1,11
24				2	CCGGTTAG	100	1.840.722	1.787.755	6603	16894	0,36	0,94
25				3	CCTCCGGT	100	2.079.973	2.071.485	47893	63597	2,3	3,07
26				4	CCTCCTTG	100	1.408.784	1.390.268	65407	71192	4,64	5,12
27			Elut	4	CCTGCCAA	100	1.376.475	1.347.476	61450	67444	4,46	5,01
28	Sörens target 1		Start	1	CGAAGCTG	76	347.246	326.677	1848	4660	0,53	1,43
29				3	CGTCCTCC	76	345.801	347.828	4770	9300	1,38	2,67
30				4	CGTCTGAG	76	378.425	381.974	18812	24940	4,97	6,53
31				5	CTAGTTGC	76	435.839	428.503	58493	72733	13,42	16,97
32			End rnd	6	CTGCAATG	76	355.856	349.518	62596	75344	17,59	21,56
33				7	CTGCTGAA	76	561.194	550.698	133206	157601	23,74	28,62
34				8	CTGGATAA	76	525.082	503.704	162315	182925	30,91	36,32
35				9	CTTGGCCT	76	502.224	486.944	169806	192166	33,81	39,46
36				10	GAACCGTT	76	317.038	307.159	124549	135021	39,29	43,96

Table 6.1 continued on next page

Table 6.1 continued

Sample				Tag sequence	Pool (nt)	Forward read	Reverse read	Used in analysis		% of reads used	
								For	Rev	For	Rev
37			2	GAACGCTG	76	299.104	285.497	1660	3731	0,55	1,31
38			3	GAATCATG	76	491.899	475.980	2599	6687	0,53	1,4
39			4	GATGGACC	76	222.241	208.059	2167	4851	0,98	2,33
40			5	GATTCCAG	76	405.535	422.321	15722	20899	3,88	4,95
41	Sörens target 2		6	GCATTGGT	76	424.129	432.667	76728	86515	18,09	20
42		End rnd	7	GCCATGCC	76	304.279	290.145	83371	88677	27,4	30,56
43			8	GCTTGATG	76	486.752	488.432	277400	292706	56,99	59,93
44			9	GGCAGCCG	76	67.572	53.781	27143	20738	40,17	38,56
45			10	GGCTTGAA	76	329.567	338.577	202698	206295	61,5	60,93
46			11	GGTAACGC	76	164.090	147.767	97545	91297	59,45	61,78
47			12	GGTCGGCG	76	207.560	213.174	113768	123197	54,81	57,79
48			13	GGTTCAGC	76	361.445	361.865	240894	252782	66,65	69,86
49			1	GTATCTGA	76	421.927	386.178	1887	4912	0,45	1,27
50			2	GTATTCTC	76	458.182	413.134	2899	6623	0,63	1,6
51			3	GTCATCTA	76	395.176	345.992	2439	5173	0,62	1,5
52			4	GTCTTGGC	76	450.350	416.149	2693	5143	0,6	1,24
53			5	GTTAAGTT	76	521.085	464.094	6535	9941	1,25	2,14
54	Sörens target 3		6	TAACGTCG	76	630.440	651.833	17881	27950	2,84	4,29
55		End rnd	7	TCAGTTAA	76	718.189	758.279	29760	46263	4,14	6,1
56			8	TCGGTCAT	76	510.011	549.094	26848	40270	5,26	7,33
57			9	TCGTTCGA	76	518.197	545.801	43235	60089	8,34	11,01
58			10	TCTTATAT	76	590.943	667.692	50996	73218	8,63	10,97
59			11	TGCGGACT	76	460.134	458.023	62823	84224	13,65	18,39
60			12	TGGACCAT	76	606.732	627.486	110205	146180	18,16	23,3
61			13	TGGCGTTA	76	386.141	392.634	52738	70222	13,66	17,88
62			14	TGGTAATT	76	464.450	507.065	47980	65569	10,33	12,93
63	Control 1 ‡			TGGTTGAC	100	2.998.366	3.028.812	2255501	2722832	75,22	89,9
64	Control 2 §			TTAGGATG	76	654.441	653.731	478699	582387	73,15	89,09
65			8	TTATCGTC	76	482.902	496.709	282350	362336	58,47	72,95
66			9†	TTCTGAAC	76	587.991	588.874	328907	412315	55,94	70,02
67	Stoltenburg target 1	End rnd	10†	TTGAGACC	76	435.779	428.292	216286	271648	49,63	63,43
68			11≠	TTGATCCA	76	542.430	529.515	279369	336267	51,5	63,5
69			12≠	TTGGCAGA	76	484.700	482.071	246744	307636	50,91	63,82
70			13ƨ	TTGGCTCC	76	380.577	375.618	184456	234635	48,47	62,47
					Unassigned	15.878.191	16.401.130				
								Total used:	49863130		

† = Negative selection step (StrepBeads)

‡ = acagatcgagggcagcgatagctgggctaataaggttagccccatcggtcctggacttgggact

§ = ggagccgaactgtctgagtagtgttgacattcttctacgt

≠ = Negative selection step (EA-TosylBeads) + subtraction Step (EA-TosylBeads) after elution EA = ethanolamine

ƨ = Negative selection step (EA-TosylBeads)

Chapter 7

A capture approach for supercoiled plasmid DNA using a triplex-forming oligonucleotide

Vincent J.B. Ruigrok, Edze R. Westra, Stan J.J. Brouns, Christophe Escudé, Hauke Smidt, John van der Oost

Nucleic Acids Research, 2013. DOI: 10.1093/nar/gkt239

Abstract

Proteins that recognize and bind specific sites in DNA are essential for regulation of numerous biological functions. Such proteins often require a negative supercoiled DNA topology in order to function correctly. In current research, short linear DNA is often used to study DNA-protein interactions. Although linear DNA can easily be modified, for capture on a surface, its relaxed topology does not accurately resemble the natural situation in which DNA is generally negatively supercoiled. Moreover, specific binding sequences are flanked by large stretches of non-target sequence *in vivo*. Here we present a straightforward method for capturing negatively supercoiled plasmid DNA on a streptavidin surface. It relies on the formation of a temporary parallel triplex, using a triple helix forming oligonucleotide containing Locked Nucleic Acid (LNA) nucleotides. All materials required for this method are commercially available. Lac repressor binding to its operator was used as model system. Although the dissociation constants for both the linear and plasmid-based operator are in the range of 4 nM, the association and dissociation rates of Lac repressor binding to the plasmid-based operator are ~18 times slower than on a linear fragment. This difference underscores the importance of using a physiologically relevant DNA topology for studying DNA-protein interactions.

Introduction

Proteins that recognize and bind specific sites in DNA are essential for controlling a wide range of biological functions at the level of DNA replication [65, 66], regulation of gene expression [267], homologous recombination [97] and various other processes. In turn, proteins involved in such processes often require a negative supercoiled (nSC) DNA topology in order to function correctly [4]. Recently, it was also shown that an nSC DNA topology is also required for the specific DNA binding of Cascade, a protein complex involved in the prokaryotic CRISPR-Cas immune system [300]. Given the importance of DNA topology, it is not surprising that a lot of effort is made to maintain a correct DNA topology *in vivo* [13, 295].

Proteins that bind specific sites in DNA also face the challenge of finding their specific binding site amongst megabases of non-target DNA. A combination of 1D diffusion (sliding) along the DNA and 3D diffusion (hopping) in the cytoplasm [85] can lead to more rapid targeting, according to the facilitated diffusion model [11, 12, 237]. Intersegmental transfer can also play a role, however, this is only relevant for proteins containing two distinct DNA binding sites such as for instance the Lac repressor and Cre recombinase [84, 249].

DNA-protein interactions can be studied using a variety of techniques, amongst others single-molecule techniques such as total internal reflection fluorescent microscopy (TIRF) [277] and surface plasmon resonance (SPR) [164]. Especially in SPR experiments, short linear target DNA is often used to study the kinetics of DNA-protein interactions. Short linear target DNA is convenient for SPR analysis, because the 3′ or 5′ ends are easily biotinylated, which allows for stable capturing on a streptavidin surface. However, such linear targets do not accurately mimic the natural situation, in which an nSC DNA topology prevails and where non-target DNA is much more abundant than specific binding sites. This might give rise to distortions in the data.

In that respect, attaching nSC plasmid DNA, containing a specific binding site, would be more appropriate to use in SPR experiments. The lack of 3′ or 5′ ends, however, makes it not straightforward to attach plasmid DNA to a surface. In the present study, we aimed to create an irreversible topological link between an nSC plasmid and a biotinylated triplex-forming oligonucleotide (TFO), which results in a padlock-modified plasmid, or catenane, that can be captured on a surface. The production of padlock-modified plasmids has previously been described for sequence specific labelling double-stranded DNA [59, 72, 242], in order to form such a complex with a good yield, a stable triple helix must be formed.

DNA triplex formation requires stretches of homo-purines (A, G) in one strand and homo-pyrimidines (C, T) on the opposite strand of the double stranded target DNA [56]. Generally, two classes of triplexes can be distinguished, according to the orientation and composition of the third strand: pyrimidine-rich third strands bind parallel to the purine strand of the duplex and form T·AT and C^+·GC triplets; alternatively, purine-rich

third strands bind antiparallel to the purine strand of the duplex and form G·GC, A·AT and T·AT triplets (in A·BC, BC indicates the natural base pair and A the third strand) [81]. Parallel triplexes only form at low pH, because protonation of the third strand cytosine (C^+) is required, whereas formation of anti-parallel triplexes is pH independent. Previous studies involving padlock-modified plasmids have mostly relied on the formation of very stable antiparallel triplexes, formed in the presence of a DNA intercalator that is not commercially available [242]. However, incorporation of locked nucleic acids (LNAs) in the pyrimidine third strand of parallel triplexes improves triplex stability and can alleviate the requirement for a low pH to some extent [274].

Here, we present a facile method for capturing of plasmid DNA on a streptavidin surface. A DNA triplex is formed by addition of an LNA-modified pyrimidine-rich biotinylated TFO, which is subsequently self-ligated in order to create a padlock-modified plasmid, or catenane. The Lac repressor has been mutated to exist as a dimer (not a tetramer) that interacts with only one DNA binding site (operator). Plasmids with and without specific Lac repressor operator sequences are used as a model to demonstrate the relevance of this approach in SPR experiments. We observed different binding kinetics to the supercoiled plasmid-based operator compared with a short linear operator. This approach therefore represents a helpful tool to study protein-DNA interactions using a DNA substrate with a physiologically relevant topology.

Results

Strategy and design

In the present study we aimed to develop a method for capturing plasmid DNA, which would allow the use of DNA with a physiologically relevant topology in SPR experiments, and in other experiments that require target immobilization. To allow the approach to be generally applicable, it should be straightforward and easily achievable with general molecular biological techniques and commercially available reagents. A general outline of the method is shown in Figure 7.1A. A newly designed target site for triple helix formation was inserted in our plasmid of interest. We have adopted a parallel triplex design (Figure 7.1B) for two main reasons. First, it does not require stabilizing molecules that are not commercially available [242]; for stabilization of the triplex, and to alleviate the requirement for a very low pH for triplex formation, the triplex-forming part of the TFO contains alternating LNA Thymine and DNA Cytosine [274]. Second, and inherently related to a parallel triplex, artifacts arising from the presence of a triplex are possibly avoided, at least at neutral pH, because parallel triplexes only form at low pH and are generally expected to dissociate at neutral pH. After addition of the closing probe, the TFO is circularized and a topological bond is introduced between the circular TFO and the plasmid. At this stage, the triple helical structure could be disrupted, as it is not required for plasmid capture anymore. In additional experiments, padlock-modified plasmids were digested with DraI. The

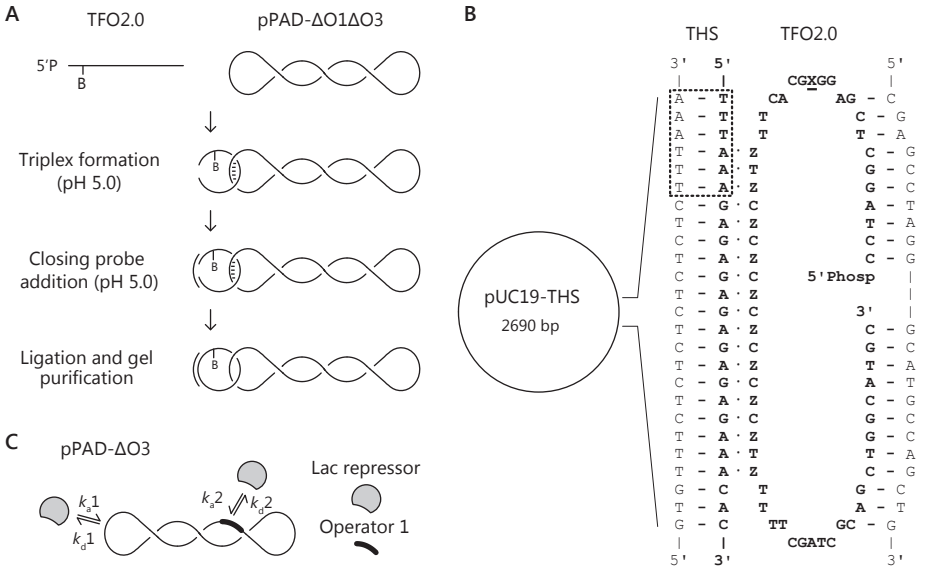


Figure 7.1: Padlock-modified plasmids. A) General procedure to prepare a padlock-modified plasmid, see materials and methods for details. The triplex (dashed lines) only exists at low pH and dissociates at neutral pH. B indicates a biotin dT, 5'P indicates a 5' Phosphate. B) The Triple Helix Site (THS) that is inserted in the vector forms a parallel triplex together with TFO2.0 . Regular Watson-Crick base pairs are indicated by "-", triplex bonds (Hoogsteen interactions) by "·". The DraI site is highlighted by a box, "z" indicates LNA thymine, "x" indicates biotin dT. C) Two interactions take place on pPAD-ΔO3, a non-specific interaction with plasmid DNA (marked by k_a1 and k_d1), and a specific interaction with the operator sequence (marked by k_a2 and k_d2).

plasmids contain two DraI sites, one of which overlaps with the THS (Figure 7.1B). One of these sites is therefore not accessible if a triplex is formed, and hence the plasmid will only be linearized. The results suggest that in 40–50% of the analyzed padlock-modified plasmids the triplex is not fully dissociated at neutral pH (Supplementary Methods and Figure S7.1), most likely due to the length of the triplex and because the TFO is heavily modified. The THS is located between the ampicillin resistance gene and the origin of replication; hence, is it located at great distance of the operator sequence, and therefore we expect the triplex not to intervene with Lac repressor binding. However, the triplex could be shortened, should the presence of the triplex intervene with the study of other binding events.

Binding affinity of Lac repressor

After padlock-modified plasmids were captured on a SA SPR chip, we performed subsequent measurements showing that Lac repressor specifically interacts with these immobilized plasmids. This confirms that the production and capture of padlock-modified plasmids was successful. Double referenced SPR data of Lac repressor binding to both padlock-modified plasmids was fitted with a simple 1:1 binding model (Figure 7.2A), yielding an overall dissociation constant (K_D) of 337 nM

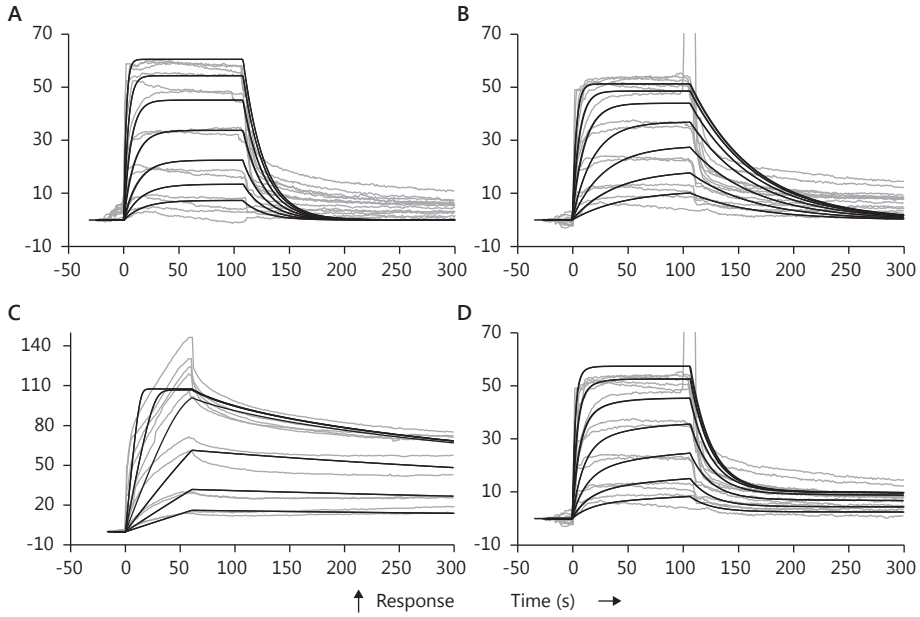


Figure 7.2: SPR data and fits. Double referenced data of replicate injections (gray) and fits to it (black). A) 1:1 fit to pPAD-ΔO1ΔO3, B) 1:1 fit to pPAD-ΔO3, C) 1:1 fit with mass transfer limitation to linear O1 DNA, D) Heterogeneous fit to pPAD-ΔO3, assuming fixed values for k_a1 and k_d1. Resulting values for the dissociation constants and kinetic rate constants are summarized in Table 7.1.

Table 7.1: Kinetic parameters for the interaction of Lac repressor with the various DNA targets. Including those found by Bondeson *et al*. [19]. Values in bold refer to interactions with operator DNA, Values in italics indicate those that were fixed during fitting with the heterogeneous model.

Target DNA	SPR							MST	EMSA
	Fit model	k_a1 ($\times 10^5$ M^{-1} s^{-1})	k_d1 ($\times 10^{-2}$ s^{-1})	K_D1 (nM)	k_a2 ($\times 10^4$ M^{-1} s^{-1})	k_d2 ($\times 10^{-4}$ s^{-1})	K_D2 (nM)	K_D (nM)	K_D (nM)
Linear O1 DNA	Mass transfer	**12**	**0.48**	**3.9**				**4.5**	
pPAD-ΔO1ΔO3	1:1	1.55	5.2	337				147	
pPAD-ΔO3	1:1	1.09	1.7	155				188	
pPAD-ΔO3	Heterogen	*1.55*	*5.2*	*337*	**6.8**	**2.7**	**4.0**		
Linear O1 DNA	n.a.	**18**	**0.034**	**0.2**					**4.2**

for pPAD-ΔO1ΔO3 (not containing an operator) and 155 nM for pPAD-ΔO3 (containing operator 1) (Figure 7.2B). Fits shown in Figure 7.2A nicely follow the association and equilibrium phase, fits shown in Figure 7.2B are not matching the data as nicely, although this is to be expected (discussed later in the text). In addition, microscale thermophoresis (MST) measurements, using the same plasmids, yield dissociation constants in the same range as SPR measurements; 147 nM for pPAD-ΔO1ΔO3, and 188 nM for pPAD-ΔO3 (Table 7.1, Figure S7.2A and B).

Double referenced data of Lac repressor binding to linear O1 DNA was fitted with a 1:1 model that takes mass transport limitation into account [192] and resulted in

a dissociation constant of 3.9 nM (Figure 7.2C), whereas a dissociation constant of 4.5 nM was obtained using MST (Table 7.1, Figure S7.2C). The association rate constant (k_a) as well as the dissociation constant (K_D) are well in range with values determined earlier [19] (Table 7.1). Despite the high flow rate used during SPR measurements, mass transport effects are to be expected because of the high association rate of Lac repressor [85]; together with the somewhat high ligand density, this feature could account for the deviation between fit and actual data. Affinity data, together with all association rate constants (k_a) and dissociation rate constants (k_d), are presented in Table 7.1. These values show that different methods yield similar affinities and that there are differences in the affinity of Lac repressor binding to both padlock-modified plasmids; this is to be expected because one of these plasmids contains a specific binding site.

Heterogeneous binding kinetics of Lac repressor

Kinetic parameters for the interaction of Lac repressor with the plasmid-based operator, were derived by fitting the binding curves of pPAD-ΔO3 with a model that assumes a heterogeneous ligand (plasmid), and that the analyte (A, Lac repressor) can bind independently to two ligand sites, specifically to the operator (C) and non-specifically to the rest of the plasmid (B) (Figure 7.1C):

Equation 7.1A $$A + B \underset{k_d1}{\overset{k_a1}{\rightleftarrows}} AB$$

Equation 7.1B $$A + C \underset{k_d2}{\overset{k_a2}{\rightleftarrows}} AC$$

The kinetic parameters for the non-target interaction (k_a1 = 1.55×10^5 M^{-1} s^{-1} and k_d1 = 0.052 s^{-1}) were derived from the interaction of Lac repressor with the pPAD-ΔO1ΔO3 plasmid and are considered to be similar for both plasmids. Hence these values were taken into account during fitting of the data, yielding a dissociation constant of 4.0 nM for the interaction of Lac repressor with the plasmid-based operator. Fits according to the heterogeneous model are shown in Figure 7.2D. This fit is much better than the fit with a 1:1 binding model as shown in Figure 7.2B. The association and dissociation rate constants for the specific interaction of Lac repressor with the plasmid based operator (k_a2 = 6.81×10^4 M^{-1} s^{-1} and k_d2 = 2.74×10^{-4} s^{-1}) are ~18 times slower in comparison with those obtained for the interaction with linear O1 DNA, while the dissociation constants are equal (Table 7.1).

Discussion

In this work, we present a straightforward approach to capture plasmid DNA on a streptavidin surface, and we demonstrate its usefulness by characterizing the

interaction between Lac repressor and plasmid DNA. Our strategy requires the insertion of a target site for triple helix formation in the plasmid of interest. Next to a pUC plasmid, this insertion can be performed in many other commercially available vectors, which also contain the targeted region (see Table 7.2). Further steps are also easy to implement, making our design a generally applicable method for plasmid capture.

Short linear target DNA is often used to study DNA-protein interactions. However, this does not accurately reflect the natural configuration of DNA, because DNA topology as well as non-target DNA are important factors contributing to binding affinity. Moreover, DNA topology has a direct influence on the interaction between DNA and many DNA-binding proteins, as exemplified by the supercoiling-dependent DNA binding of Cascade [300]. This is also demonstrated by the binding of DnaA to the origin of replication (*oriC*) on the *Escherichia coli* genome: this complex is more stable if *oriC* has an nSC topology [65, 66]. In addition, it has been shown that promoters can be stimulated or inhibited by increased negative supercoiling [267], most likely related to the binding efficiency of the RNA polymerase complex.

Here we have selected the *E. coli* Lac repressor as a model system, because this protein and the three operators it can bind are well studied [154, 184, 205]. It has been shown that supercoiling has an effect on the dissociation of Lac repressor-operator complexes [57, 301] and on Lac repressor-mediated DNA looping [201]; however, the proteins used in these studies were all naturally occurring tetramers.

Specifically, the interaction of Lac repressor with O1 has been studied in great detail, and remarkably high association rates have been reported based on equilibrium methods, spanning the range between 1×10^{8} –1×10^{10} M^{-1} s^{-1} [76, 238]. It is, however, more relevant to compare our results to those obtained in a previous SPR analysis of this interaction [19], in which the following kinetic parameters were determined, a k_a of 1.8×10^{6} M^{-1} s^{-1}, a k_d of 3.4×10^{-4} s^{-1} and a K_D of 0.2 nM. In the latter study, the dissociation constant was also determined using an electrophoretic mobility shift assay: a K_D of 4.2 nM. This number is well in range with the values we find, using SPR (3.9 nM and 4.0 nM) and MST (4.5 nM). The major difference between both studies is that Bondeson *et al.* [19] used a wild type (tetrameric) Lac repressor, which can bind two operators at the same time, whereas we used a mutated (dimeric) Lac repressor that can bind only one operator at the same time. In general, the presence of multiple binding sites is disadvantageous for SPR analysis; it could give rise to avidity effects and substantial rebinding, which results in a higher apparent affinity [189]. This explains why our dissociation rate from the linear O1 DNA (4.8×10^{-3} s^{-1}) is substantially higher, and hence results in a lower affinity (3.9 nM).

The affinities for interaction with the captured plasmids are considerably lower when compared with the linear target DNA, but Lac repressor still binds with nanomolar affinity to the plasmids. Binding curves of both captured plasmids were

initially fitted to a 1:1 binding model. A comparison of these 1:1 fits already shows differences between both plasmids; association rates are in the same range, but the dissociation from the pPAD-ΔO3 plasmid appears to be three times slower. The higher affinity interaction with the pPAD-ΔO3 plasmid, is in line with the fact that a specific binding site (operator 1) is present on this plasmid. Affinity values obtained using MST are in the same nanomolar range. They do not follow the trend that the affinity for pPAD-ΔO3 is higher than for pPAD-ΔO1ΔO3. For MST to be accurate, it is essential that the DNA concentrations are precisely known, however, in the course of this project it has proven difficult to accurately measure concentrations of highly concentrated, and hence viscous, DNA preparations. We believe this to be the origin of the discrepancy between these values.

Although previously fitted with a 1:1 binding model, binding of Lac repressor to the pPAD-ΔO3 plasmid should actually be considered as a heterogeneous event. Lac repressor can independently bind to either non-target DNA or operator DNA. We fitted the pPAD-ΔO3 binding data with a model for heterogeneous binding, in order to obtain the kinetic parameters for the secondary, specific, interaction. To do so, we assumed the kinetic parameters for the non-target interaction to be similar for both plasmids and used these as known variables for k_a1 and k_d1. As such we found a K_D2 of 4.0 nM for the interaction between Lac repressor and its plasmid-based operator. This is remarkably close to the values we found using SPR (3.9 nM) and MST (4.5 nM) (Table 7.1). Interestingly, the actual kinetics are widely different for the interactions of Lac repressor with the plasmid operator and linear O1 DNA. Both the association and dissociation rates are ~18 times slower for binding to the plasmid operator, indicating that the presence of negative supercoiling and non-target DNA indeed has a considerable effect on the actual kinetics of binding.

Conclusion

In the work presented here, we demonstrate the feasibility and usefulness of a newly developed plasmid capture approach, by applying it for the characterization of Lac repressor binding. To our knowledge, this is the first time that SPR has been used to determine the affinity and kinetic parameters of the interaction between a protein and its specific target sequence, which is located on a supercoiled plasmid. We believe this to be a versatile approach that could be useful in SPR, single molecule and other experiments, to expand the range of substrates for DNA-protein interactions beyond the use of short linear target DNA. In addition, the biotin in TFO2.0 could be replaced by other functionalities, such as fluorophores, and thus will enable studies requiring plasmid visualization. The use of padlock-modified plasmids provides a useful addition to the molecular biology toolbox, and may be used to uncover properties of supercoiling dependent proteins, that could not be studied before.

Materials and methods

Oligonucleotides

All oligonucleotides used in this study, except for TFO2.0, were obtained from Sigma and ordered without any special requirements. TFO2.0 was ordered from Eurogentec and PAGE-purified by the manufacturer. Details on all oligonucleotides and their sequences are given in Table 7.2.

Table 7.2: Oligo sequences used in this study.

	Purpose	Modification	
BG3534†	THS insertion		ctttttctacggggtctgacg***tttaaa***gagagagagagagaaacacgttaagggattttggtca
BG3535†	THS insertion		tgaccaaaatcccttaacgtg***tttaaa***cgtcagaccccgtagaaaag
	Targeted region		ctttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtca
BG3554	THS sequencing		gctgaagccagttaccttcg
	TFO2.0 Eurogentech (PAGE purified)	5′-phosphate X=Biotin dT, Z=LNA T	cctaggctcgaggxgcacttztzczczczczczczczttttcgatcgcagctggcatgc
BG3812	Closing probe		cgagcctagggcatgccagctg
BG3874	Cloning Lac repressor		gatccatgggcaaaccagtaacgttatacgatgtcg
BG3917	Cloning Lac repressor	HRV3C site and an 8x His-tag	ggtgagctcttagtggtggtgatggtgatgatggtgagcggagggtccctgaaagaggac ttcaagcgcccagggtggtttttc
BG3962	ΔO1ΔO3		aagccatggtcacacaggaaacagctatgac
BG3963	ΔO3 / ΔO1ΔO3		aagccatggcgctttccagtcgggaa
BG3964	ΔO3		aagccatggaatgtgagttagctcactcattagg
BG4162	Linear O1 S	5′-biotin	tgtgtggaattgtgagcggataacaatttcacaca
BG4163	Linear O1 A		tgtgtgaaattgttatccgctcacaattccacaca

† Underlined is the Triple Helix Site, the DraI site in Italics.

Cloning of Lac repressor coding sequence

The first 331 residues of the coding sequence of *Lac repressor*, hence excluding the C-terminal tetramerization domain (residues 340–357), were PCR amplified from a pCDF-1b plasmid. The coding sequence was amplified in 3×100 µl PCR, containing 15 ng of template DNA, 0.2 mM of each dNTP, 0.2 mM of primers BG3874 and BG3917 (containing a HRV3C site and an 8×His-tag), 1×Buffer HF and 3 Units Phusion II (Finnzymes). The PCR program was as follows: 30 s at 98°C, 5 cycles of 10 s at 98°C, 20 s at 60°C, 60 s at 72°C, 25 cycles of 10 s at 98°C, 20 s at 70°C, 60 s at 72°C, followed by 5 min at 72°C after the last cycle. Purified PCR product and destination vector, pWUR533, were subsequently digested with NcoI and SacI restriction enzymes. Both fragments were ligated together and the resulting plasmid (pWUR533-LacI) was transformed to the *Escherichia coli* strain XL1-blue for plasmid propagation.

Expression of recombinant Lac repressor

The pWUR533-LacI plasmid was transformed to *E. coli* BL21(DE3) pSJS1244 for protein expression. Fresh LB medium, containing ampicillin (final concentration: 100 µg ml^{-1}) and spectinomycin (final concentration: 50 µg ml^{-1}), was inoculated with 1% overnight

culture. Protein expression was induced three hours later by adding isopropyl-β-D-thiogalactopyranoside (IPTG), to a final concentration of 1 mM. After another three hours, cells were harvested by centrifugation. Cell pellets were either processed immediately, or stored at -20°C until further processing.

Purification of recombinant Lac repressor

Cell pellets were resuspended in 150 mM NaCl and 10 mM HEPES (pH 7.4) and subsequently lysed using a French press. Clarified cell free extract was passed through a 0.45 µm filter before incubation with HIS-Select® Nickel affinity gel (Sigma-Aldrich) which was equilibrated with water (3 ml) and 3 column volumes wash buffer [50 mM sodium phosphate buffer (pH 8.0) with 0.3 M sodium chloride and 5 mM imidazole]. After incubation, the column material was washed three times with four column volumes wash buffer. Bound protein was eluted in fractions of 2 column volumes using elution buffer [50 mM sodium phosphate buffer (pH 8.0) with 0.3 M sodium chloride and 250 mM imidazole]. Eluted protein was further purified on a Superdex 200 column (GE Heathcare), using 150 mM NaCl and 10 mM HEPES (pH 7.4) as running buffer; peak fractions corresponding to the dimer protein were collected and used for SPR and MST measurements.

Insertion of a Triple Helix Site in pUC19

A purine rich Triple Helix Site (THS, 19 bp), required for triple helix formation, was inserted between the ampicillin resistance gene and the origin of replication in a pUC19 vector. To do so, 25 ng plasmid was amplified in 50 µl using 0.2 mM dNTPs each, 125 ng of primers BG3534 and BG3535, 1×Buffer HF and 1 Unit Phusion II (Finnzymes). The PCR program was as follows: 90 s at 98°C, 18 cycles of 10 s at 98°C, 30 s at 58°C, 80 s at 72°C, followed by 7 min at 72°C after the last cycle. After PCR, 10 units of DpnI (Fermentas) was added directly to the PCR reaction mixture, for degradation of template DNA, and left at 37°C for 2 h. The resulting plasmid (pUC19-THS) was transformed to chemical competent *E. coli* DH5α. Insertion of the THS was confirmed by restriction analysis (using DraI) and sequencing (GATC, Constance, Germany) using primer BG3554.

Removal of lac operator sequences

The two *lac* operator sequences in pUC19-THS, operator 1 (aattgtgagcggataacaatt) and operator 3 (ggcagtgagcgcaacgcaatt), were removed to yield two plasmids. The pPAD-ΔO3 plasmid only contains operator 1, whereas pPAD-ΔO1ΔO3 contains no operator at all. The operator sequences were removed by PCR amplification of the whole plasmid, except for the operator region, using primers BG3963 and BG3964 for ΔO3, and BG3962 and BG3963 for ΔO1ΔO3. Plasmid, 35 ng, was amplified in 100 µl using 0.2 mM dNTPs each, 0.2 mM of each primer, 1×Buffer HF and 1 Unit

Phusion II (Finnzymes). The PCR program was as follows: 60 s at 98°C, 5 cycles of 10 s at 98°C, 10 s at 60°C, 1 min at 72°C, followed by 25 cycles of 10 s at 98°C, 10 s at 70°C, 1 min at 72°C followed by 7 min at 72°C after the last cycle. Purified PCR product was digested with 10 units of DpnI and NcoI and left at 37°C for 2 h. Fragments were ligated and transformed to electro competent *E. coli* XL1-blue cells. Successful removal of the operator sequences was confirmed by sequencing (GATC, Constance, Germany) using the standard M13-F primer.

Padlock formation

For padlock formation, approximately 50 nM plasmid (use of more plasmid should be prevented, as this could result in lower yields because of molecular crowding) was mixed with 1 µM TFO2.0, acid buffer (20 mM $MgCl_2$ and 20 mM ammonium acetate pH 5) was added to a total volume of 10 µl; low pH is needed to protonate cytosine on the third strand [72, 81], which is required in this approach. Plasmid and TFO were heated to 80°C and cooled down to 20°C at a rate of -1°C min^{-1} in a G-Storm GS1 thermocycler (start at 80°C for 30 s, then step-wise decrease to 20°C, in 350 subsequent steps in which the temperature drops with 0.17°C and stays stable for 8 s). When at 20°C, 2 µM closing probe (BG3812) was added, followed by 1.5 µl 10×T4 DNA ligase buffer (Fermentas), 1 µl of a 5 mM ATP solution and 5 Units (1 µl) of T4 DNA ligase (Fermentas). The mixture was incubated overnight at room temperature. Padlock-modified plasmids were purified from a 0.8% agarose gel (Fermentas Kit) in order to remove excess TFO and closing probe, before capturing them on a Biacore SA chip. To yield enough padlock-modified plasmid for capture, padlock formation was routinely performed in three parallel reactions, only to be pooled during the gel purification procedure.

Surface Plasmon Resonance measurements

Experiments were performed in a Biacore 3000 system (BIACORE, Uppsala, Sweden) at a constant temperature of 25°C, using 10 mM HEPES at pH 7.4, and 150 mM NaCl as running buffer. Padlock-modified plasmids were captured on an SA chip at a flow of 5 µl min^{-1}, pPAD-ΔO3 to a response of 140 RU, and pPAD-ΔO1ΔO3 to a response of 80 RU. An empty channel served as reference surface. For the padlock-modified plasmids, kinetic measurements were performed by injecting 20 µl Lac repressor at a flow of 10 µl min^{-1}, using the kinject command, followed by a 5 min dissociation phase. In total, seven concentrations (2650, 883, 294, 98, 33, 11 and 4 nM) were injected twice.

Linear Operator 1 (O1) DNA, prepared by hybridizing BG4162 and BG4163, was captured on another SA chip to 334 RU. Kinetic measurements were performed by injecting 90 µl Lac repressor at a flow of 90 µl min^{-1}, using the kinject command, followed by a 5 min dissociation phase. An empty channel served as reference

surface. In total, six concentrations (577, 192, 64, 21, 7 and 2 nM) were injected twice. Data were processed using Scrubber (BioLogic Software, Campbell, Australia), and double referenced data were analyzed using BIAevaluation software provided with the Biacore.

Microscale Thermophoresis (MST)

Plasmids for MST measurements were purified from 100 ml of overnight cultures, using a Jetstar 2.0 maxiprep kit (Genomed). The obtained pellet was re-suspended in 50 µl milli-Q water; DNA concentrations were calculated from gel, by comparing peak intensities of linearized plasmid and the 3000 bp band of a marker (1 kb, Fermentas). Purified Lac repressor was labelled using a protein labelling kit, L003 Monolith™ (NanoTemper. München, Germany). MST measurements were performed in standard capillaries on a Monolith NT.115 machine, using 5% LASER power and 40% LED power. DNA concentrations were varied, while keeping the protein concentration constant at 25 nM. Protein was diluted to 50 nM in buffer (20 mM HEPES at pH 7.4, 300 mM NaCl,) and subsequently mixed in a 1:1 ratio with dilutions of DNA (in milli-Q). Data were analyzed using the software provided with the Monolith NT.115 (NanoTemper. München, Germany).

Acknowledgements

The authors wish to thank Intan Taufik (Groningen Biomolecular Sciences and Biotechnological Institute and Molecular Microbiology, Groningen) for help with MST measurements and Mitch Lewis (University of Pennsylvania) for helpful discussions on recombinant expression of Lac repressor.

Funding

This work was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research (ALW-GO-PL/08-08) and by an NWO Vidi grant to SJJB (864.11.005).

Supplementary information

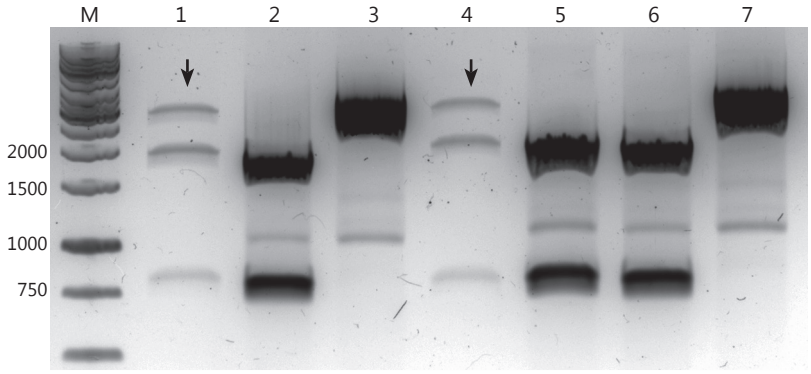
Materials and methods

Removal of two DraI sites from pPAD-ΔO1ΔO3 and pPAD-ΔO3

Two DraI sites were removed to simplify restriction analysis. To do so, 25 ng plasmid was amplified in 50 µl using 0.2 mM dNTPs each, 125 ng of primers BG4394 (cctagatccttttCaattaaaaatgaagttttCaatcaatctaaag, changed nucleotides are in capitals) and BG4395 (ctttagattgattGaaaacttcatttttaattGaaaaggatctagg), 1×Buffer HF and 1 Unit Phusion II (Finnzymes). The PCR program was as follows: 90 s at 98°C, 16 cycles of 10 s at 98°C, 30 s at 58°C, 60 s at 72°C, followed by 7 min at 72°C after the last cycle. After PCR, 10 units of DpnI (Fermentas) was added directly to the PCR reaction mixture, for degradation of template DNA, and left at 37°C for 2 h. The resulting plasmids were transformed to chemical competent *E. coli* DH5α. Removal of both DraI sites was confirmed by restriction analysis (using DraI) and sequencing (GATC, Constance, Germany) using primer BG3554.

Triplex restriction analysis

Padlocks for triplex restriction analysis were prepared as described in the main text, using five parallel reactions of each plasmid. After gel purification, padlocks were incubated with streptavidin coated magnetic beads (Invitrogen) for 45 min at RT under gentile rotation to prevent settling of the beads. Subsequently, beads were washed several times to remove unbound plasmid/padlock. Beads were used as template for restriction analysis, incubation was 45 min at 37°C for 45 min under gentile rotation to prevent settling of the beads. Enzymes were heat inactivated by incubation at 65°C for 20 min and run on gel (stained with SYBR-gold).



M.	Marker	
1.	Padlock pPAD-ΔO1ΔO3 + DraI	
2.	pPAD-ΔO1ΔO3 + DraI	Expected: 1795 + 788 bp
3.	pPAD-ΔO1ΔO3 + NcoI	Expected: 2583 bp
4.	Padlock pPAD-ΔO3 + DraI	
5.	Beads + pPAD-ΔO3 +DraI	Expected: 1902 + 788 bp
6.	pPAD-ΔO3 + DraI	Expected: 1902 + 788 bp
7.	pPAD-ΔO3 + NcoI	Expected: 2690 bp

Figure S7.1: Triplex restriction analysis. Arrows indicate linearized vector, its presence shows that one DraI restriction site is not accessible, suggesting that a triplex is still present in 40–50% of the analyzed padlock plasmids. The two lower bands are products yielded after double digestion of the vector.

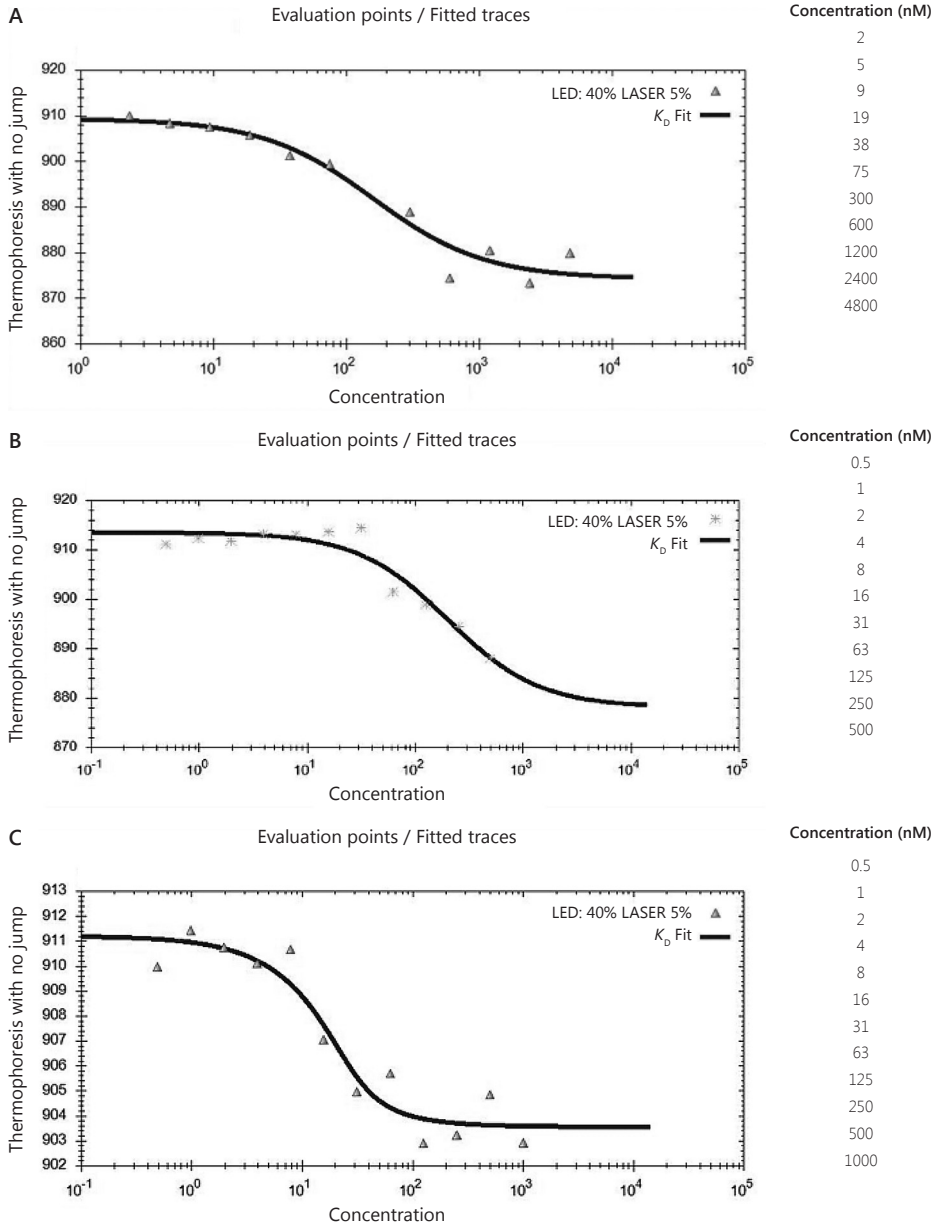


Figure S7.2: Microscale thermophoresis data. Curves fitted through the data points are shown as well. A) pPAD-ΔO1ΔO3, B) pPAD-ΔO3, C) Linear O1 DNA.

Chapter 8

Summary and general discussion

Summary

This thesis focuses on the selection and characterization of DNA aptamers and the various aspects related to their selection from large pools of randomized oligonucleotides. Aptamers are affinity tools that can specifically recognize and bind predefined target molecules; this ability, however, is not exclusively associated with aptamers. Antibodies are the most successful affinity tools used today, but alternative affinity tools such as aptamers, engineered binding proteins and molecular imprinted polymers are emerging as sound alternatives. A comparison of their properties is described in **Chapter 1**. The strength and specificity of the interaction between an affinity tool and its target molecule is an important feature. Generally, an affinity tool should have a high affinity for its target and should be highly specific in order to be useful for research or commercial purposes. One highly advanced method to characterize the interaction between an affinity tool and its target molecule makes use of a Surface Plasmon Resonance (SPR)-based biosensor. Although SPR is an optical phenomenon, in depth knowledge of the physics behind this phenomenon is not required to operate an SPR-based biosensor. Experiments should be performed in a correct way, and therefore it is important to understand how experimental parameters, such as flow rate, ligand density, surface preparation, and reagent quality either improve or adversely affect data quality. Experimental considerations, as well as methods for proper data analysis are discussed in **Chapter 2**. Data generated within the framework of the 2011 Global Label-free Interaction Benchmark study serves as a typical example.

The ability of aptamers to bind a specific target originates from an intricate interplay between the oligonucleotide sequence and the three dimensional structure that this sequence allows to form. In **Chapter 3** this is illustrated by the selection and characterization of streptavidin-binding aptamers. Five aptamer families were identified, sharing a similar secondary structure. Although slight variations at the actual sequence level are present, two guanines are completely conserved. Using site-specific mutagenesis it was demonstrated that these guanines are essential for streptavidin binding. Binding kinetics and the dissociation constant of each aptamer was determined by SPR and were all within the range of 35–375 nM. Two aptamers can bind one streptavidin tetramer at the same time, as was shown by native mass spectrometry analysis. In addition, the three dimensional structure of the most abundant aptamer was modelled and manually docked to the streptavidin structure, in order to gain more insight in the molecular basis of the interaction. To extend this knowledge even further, crystallization trails, aiming to obtain a co-crystal structure for the streptavidin-aptamer complex, were performed, and are described in **Chapter 4**. Unfortunately, these trials did only yield protein crystals, instead of the desired streptavidin-aptamer complex. Therefore, alternative experimental and

computational approaches were investigated that could be used to study aptamer-protein interactions. Combining techniques as SPR, small-angle X-ray scattering (SAXS), isothermal titration calorimetry (ITC), and Dynamic light scattering (DLS) could be considered as an alternative to X-ray crystallography. In addition, some of these techniques may provide information on the dynamics of complex formation, whereas crystallography gives a time- and position-averaged image.

Besides streptavidin, another protein, SpaC, was subjected to aptamer selection in this thesis. SpaC is a subunit of pili present on the probiotic Gram-positive bacterium *Lactobacillus rhamnosus* GG and contains a binding domain for human-mucus. Presence of this binding domain is considered an advantage, because it is already designed to interact with other molecules. Successful production and purification of recombinant SpaC protein is described in **Chapter 5**, as well as the characterization of DNA oligonucleotides enriched during subsequent selection rounds. Sequence analysis revealed that specific oligonucleotides are indeed enriched. Furthermore, results of pilot SPR experiments indicated that they bind specifically to SpaC, but more detailed experiments are required to unambiguously demonstrate this.

The dynamics of aptamer enrichment are poorly understood. To address this issue and to gain a more fundamental insight in the aptamer selection process, a multiplexed high throughput sequencing effort was started, which is described in **Chapter 6**. In this approach samples of 70 selection rounds, derived from 8 distinct aptamer selection experiments, were barcoded, pooled together and sequenced; over 84 million paired-end reads were obtained and analyzed. Samples enriched to bind streptavidin show a decrease in α-diversity across subsequent selection rounds. Interestingly, large differences were found between the composition of fractions enriched by affinity elution and thermal elution. Moreover, a small scale comparison of two clone libraries showed that affinity elution, which is expected to enrich more specific binders, also specifically enriches rapid binders.

Supportive SPR experiments have made an important contribution throughout this thesis. The main focus in **Chapter 7**, however, is on a new application of SPR. The development of a capture approach for supercoiled plasmid DNA, using a triple helix forming oligonucleotide, is described. It could be demonstrated that plasmid DNA can indeed be captured and that SPR can subsequently be used to derive kinetic parameters of a specific interaction with a plasmid. In this particular case the interaction between Lac repressor and its plasmid-based operator was characterized, showing that the association and dissociation rates are ~18 times lower, but that the affinity is the same, when compared to binding to linear operator DNA. This difference underscores the importance of using a DNA substrate with a physiologically relevant topology for studying DNA-protein interactions.

General discussion

Since the discovery of aptamers in 1990 [58, 285] the field of aptamer research has progressed rapidly and has matured to a large and diverse field of research. In the course of the research project described in this thesis, ample developments have been accomplished in aptamer research. If the current project were to be re-initiated today, some of these developments could have influenced the strategy for aptamer selection that was adopted in this work. In the remainder of this chapter these developments are discussed, as well as how they could be used to enhance the success-rate of future aptamer selection endeavors.

Target selection

Small organic compounds

A large number of aptamers are available that are capable of binding a wide variety of small organic compounds [269, 270]; however, not every small organic molecule makes a good potential target for aptamer selection. Interactions between aptamers and small organic molecules occur via a combination of aromatic stacking, hydrogen bonding, electrostatic interactions and target encapsulation [89]. Therefore, a large number of organic compounds that are targeted by aptamers have planar surfaces engaging in aromatic stacking interactions, or functional groups that allow hydrogen bonding, or most often a combination thereof. Examples of suitable aptamer targets include amino acids [60, 71], tetracycline antibiotics [197, 308], cofactors [150], hormones [136], nucleotides [110, 250], and pharmaceuticals [118]. In addition, molecules containing amino-modified sugar groups (e.g. aminoglycoside antibiotics) are also often targeted successfully [264, 298]. Hence taking into account appropriate criteria for the selection of a target molecule that contains some of the aforementioned features, could increase the chances of successful aptamer enrichment.

Other, more practical considerations are that the target molecule should be applicable for coupling to a matrix. A particular challenge here is confirming successful coupling of the target to the matrix; infrared spectroscopy has proven useful to this end [175]. In addition, X-ray Photoelectron Spectroscopy (XPS) could for instance also be used to detect changes in the elemental composition of treated and untreated surfaces. Moreover, a target for aptamer selection should preferentially be water-soluble. Although one aptamer is known that still binds its target in the presence of 20% methanol [42], it was selected in an aqueous environment without methanol.

Proteins

Similar to small organic molecules, DNA-protein interactions also occur by a combination of aromatic stacking, hydrogen bonding and electrostatic interactions; however, proteins are recognized via shape complementarity, rather than target

encapsulation. Some of the early protein-binding aptamers were selected to bind proteins that bind nucleic acids as part of their natural function, such as T4 DNA polymerase [285], HIV-1 Reverse transcriptase [286], and Taq DNA polymerase [43]. Nowadays a variety of aptamers are available that bind proteins that are not associated with nucleic acid binding at all, but it is not possible to predict beforehand which proteins will successfully be bound by aptamers. A common feature shared between these proteins, however, seems to be that they all have a binding site, either for protein-protein interactions, or for interactions with smaller molecules. Non-interacting (hydrophilic) protein surfaces might simply not provide enough support to facilitate aptamer binding. On the contrary, binding sites are designed to bind other structures, and hence might provide enough support to facilitate recognition by nucleic acid aptamers. This might explain why no obvious aptamer for hemoglobin could be enriched. Its oxygen binding pocked is most likely too small, and the protein surface might be too smooth to support binding of a nucleic acid aptamer. Presence of a binding site is on its own not a decisive factor for successful enrichment, as is supported by the ambiguous results of the enrichment of a specific aptamer for the mucus-binding protein SpaC described in this thesis (Chapter 5). Other factors, for instance protein stability, are important contributors to successful aptamer enrichment as well.

Nucleotide developments

Increased oligonucleotide stability

Throughout literature, both DNA and RNA aptamers are described, and the choice for either nucleic acid as a basis for aptamer development might seem arbitrary. This particular choice probably depends on practical considerations and partly on the final application. An obvious practical drawback of RNA is that it is prone to hydrolysis, by nucleophilic attack of the 2′-OH group on the near phosphorus atom, and enzymatic degradation. DNA lacks the 2′-OH and hence is inherently more stable. In addition, selection of RNA aptamers is more laborious as its processing requires additional enzymatic steps. On the other hand, RNA has a relatively flexible backbone in comparison with DNA, which enables it to adopt a wider range of three dimensional conformations. In that respect, RNA can potentially bind to a broader range of target molecules. Ideally, the stability of DNA is combined with the flexibility of RNA, and a solution for this is found in the use of modified nucleotides [168] (Figure 8.1). Various modifications result in increased stability, they comprise capping of the 5′ or 3′ ends, replacing the phosphate backbone by a phosphorothioate one, and modifications of the (deoxy)ribose sugar, such as locked nucleic acids; recently the use of alternative sugars has been reported, for instance to obtain arabinose nucleic acid (ANA) and anhydrohexitol nucleic acid (HNA) [219]. The modifications that are most frequently used nowadays, however, are those in which the 2′-OH has been replaced, for

instance by a Fluor atom (2′-F) or an amino group (2′-NH_2). Several of such modified nucleotides are commercially available (e.g. http://www.trilinkbiotech.com). Another factor attributing to their popularity is that several modified RNA polymerases are described that are capable of incorporating such 2′-modified nucleotides [35, 209]. In other words, these 2′-OH modified nucleotides are compatible with the enzymatic steps of the selection procedure. At the start of this project, DNA was preferred above RNA for reasons of stability and ease of selection. However, if this project were to be re-initiated, an RNA library also containing 2′-modified nucleotides could be a viable alternative to DNA libraries, given their increased availability, and providing a combination of stability and large structural diversity.

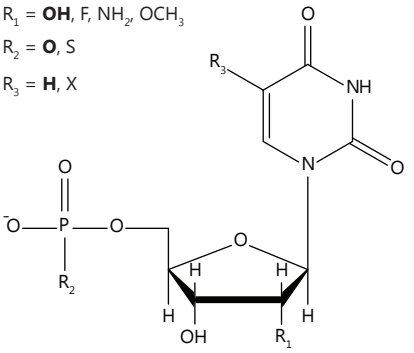


Figure 8.1: Frequent modifications on nucleosides. Modifications leading to increased stability are most often made on positions indicated by R_1 and R_2. The C5 position of uracil and thymine, R_3, are often targeted to increase diversity of the building blocks (X), e.g. [115, 145, 149]. Uridine 5′-phosphate is shown here, groups normally occurring are shown in bold.

Increased variation of building blocks

Although aptamers with affinities in the picomolar range have been described [244], the majority of aptamers have affinities in the nanomolar, or even micromolar range. Antibodies, on the contrary, often have affinities in the low picomolar range. The foundation for this difference in affinity lies in the number and variation of the building blocks. DNA or RNA aptamers generally consist of only four building blocks with limited variation, whereas antibodies consist of twenty building blocks with a variety of amino acid side chains, allowing for much more variations and types of interactions to take place. Similar to aptamer-target interactions, antibody-antigen interactions occur through electrostatic interactions and hydrogen bonding. In addition, however, antibodies also interact through hydrophobic interactions with their antigen. Hydrophobic interactions are important, as their strength is proportional to the surface involved, and for some antibody-antigen interactions hydrophobic interactions account for most of the binding energy [116]. The four natural bases are not able to form hydrophobic interactions, resulting in a physical drawback for aptamers. In fact, the limited number of aptamer building blocks and their similarity are the biggest limitation of aptamers.

A variety of modified nucleotides have been described that increase the chemical diversity and properties of aptamer building blocks. Several modifications are

fully compatible with the enzymatic steps of the selection procedure and allow for (exponential) production by PCR from modified templates [168]. Other modifications are not fully PCR-compatible but they can still be incorporated by primer extension methods, allowing (linear) production from unmodified templates. Some other modifications, however, can only be applied to aptamers after the selection procedure, for instance addition of a 3′ or 5′ functionality. Such modifications are referred to as post-selection modifications. Another interesting example, although not necessarily post-selection, is the attachment of a mixture of compounds, by dynamic combinatorial chemistry, to the 2′-NH_2 group of amino-modified ribose rings [25]. The C5-position on thymine and uracil nucleobases (R_3 in Figure 8.1) are the most suitable sites for modifications, because modifications at these sites are most often still accepted by polymerases [130]. Hence it is not without reason that most modifications that increase the chemical variation of aptamer building blocks target this C5-position [115, 145, 149]. Diversified oligonucleotides, however, do not necessarily yield higher affinity binders as exemplified by a modified thrombin aptamer that has a 4–10 fold weaker affinity as compared to the unmodified sequence [149].

A more recent development concerns the modification of nucleotides that enable hydrophobic interactions between aptamer and target [290]. Also in this case, uracil bases were modified at their C5-position, by the addition of various aromatic and aliphatic hydrophobic groups. Although the modifications were not fully compatible with PCR, primer extension could be used to produce sufficient material for aptamer selection. Indeed DNA aptamers containing modified nucleotides were selected (with K_D values of 4 and 6 nM) to a protein target (tumor necrosis factor receptor superfamily member 9, TNFRSF9) for which enrichment of unmodified DNA aptamers remained unsuccessful in the past. Hydrophobic interactions are a key requirement for this interaction; recently the importance of hydrophobic interactions was also demonstrated for another target, by the description of a crystal structure of such a modified aptamer and platelet-derived growth factor B [47].

Altogether, modified nucleotides greatly increase the chemical variation of aptamer building blocks, which are now no longer constrained to the limited properties of natural nucleic acids. Although these developments broaden the range of potential targets, the limited availability and the required chemical knowledge on the synthesis of such modified nucleotides, poses a challenge for their widespread availability in the near future.

Commercialization of aptamers

Therapeutics

The most profound commercial application of aptamers today clearly demonstrates their large commercial potential. At the same time it also highlights a limitation that, amongst other reasons, prevents commercialization at a large scale.

Macugen (Pegaptanib) became the first FDA-approved aptamer therapeutic in 2004 [82], and remains the only one today. It is approved for the treatment of age-related macular degeneration [196], because the RNA aptamer can bind all isoforms of vascular endothelial growth factor (VEGF), except for the smallest one ($VEGF_{121}$), and prevents subsequent receptor-binding [244]. Already two years later, an antibody fragment-based therapeutic (Lucentis, Genentech) was approved, also for treatment of age-related macular degeneration, and it quickly took over the market, because it binds all isoforms of VEGF, including the smallest. Therefore Macugen's approval not only marks a milestone in aptamer research and commercialization, it also shows the supremacy of antibodies and antibody-based therapeutics. However, many aptamer-based therapeutics are still in clinical trials, and hold the promise of a new generation of aptamer-based therapeutics in the near future [131, 275].

Currently, Altermune Technologies LLC (Irvine, CA, USA) is developing a new and exciting application of aptamers that could result in new therapeutics and alternatives for antibiotics. In their approach, an aptamer that is specific for a particular non-immunogenic target, e.g. a pathogen or a cancer cell, is fused to the α-gal epitope, to obtain the Altermune linker. The α-gal epitope is known to elicit a strong immune response, therefore the non-immunogenic target becomes highly immunogenic upon binding of the Altermune linker. As a result the target will be neutralized by the subsequent immune response [186]. The principle of this chemical programming of immunity has already been shown to be effective with small organic molecules [174, 221]. According to a recent global market analysis, the market of aptamer-based therapeutics is expected to grow from $13 million in 2012 to a value of $1.7 billion in 2017 [114].

Diagnostics

Another area in which aptamers have potential applications is that of diagnostics. Over the years many aptamers have become available for a variety of targets. The use of aptamers in various assay-formats using various detection methods have been described in academic journals, but none of these were successfully commercialized. For any new technology to become successful, it must provide a clear advantage over any existing technology. Therefore, the widespread use of antibodies and their well-established methods poses a major roadblock for successful introduction for aptamer-based diagnostics. It was only recently that the first aptamer-based diagnostic test became commercially available [75], for the detection of Ochratoxin A in various food matrixes (NeoVentures Biotechnology Inc, London, Canada). Although the use of aptamers is not as strongly rooted in current research, this aptamer-based assay demonstrates that there is indeed a commercial potential for dedicated aptamer-based diagnostics.

Recently, SomaLogic (Boulder, CO, USA) introduced a new aptamer-based proteomics method [77]. Their method can be used to simultaneously identify and

quantify 813 proteins. This method is a useful diagnostic tool that, among others, could be used to identify and validate potential drug targets. This breakthrough is enabled by the development of slow off-rate modified aptamers (SOMAmer). According to the same market analysis, the market for aptamer-based *in vitro* diagnostics is expected to grow from an expected value of $86 million in 2012 to $2 billion in 2017 [114], exceeding that of aptamer-based therapeutics.

Custom aptamer development

A convenient aspect of using antibodies for research applications is that antibodies for numerous targets are readily available. If an antibody for a specific target is not available, an array of polyclonal antibodies could quite easily be generated by immunizing a mouse or rabbit and subsequent (crude) antibody purification from blood plasma. Numerous companies offer services for the generation of such custom polyclonal antibodies. Aptamers, on the contrary, are not so readily available. Even if a target specific aptamer sequence is available, it should be confirmed that this sequence is indeed target specific and is not binding to the matrix used during selection. However, more and more companies are emerging that offer custom aptamer selection, probably using automated SELEX procedures, analogous to generation of custom polyclonal antibodies. Whether this coincides with the expiry of the original patents on aptamer technology is not completely clear, but it signals a new stage in the application and development of aptamers.

Appendices

References

1. (2008). *Biacore Sensor Surface Handbook*. Uppsala: GE Healthcare Bio-Sciences AB.
2. **Abdiche YN, Malashock DS, Pons J.** (2008). Probing the binding mechanism and affinity of tanezumab, a recombinant humanized anti-NGF monoclonal antibody, using a repertoire of biosensors. *Protein Sci.* 17:1326-35
3. **Baker BR, Lai RY, Wood MS, Doctor EH, Heeger AJ, Plaxco KW.** (2006). An electronic, aptamer-based small-molecule sensor for the rapid, label-free detection of cocaine in adulterated samples and biological fluids. *J. Am. Chem. Soc*. 128:3138-9
4. **Bates AD, Maxwell A.** (2005). *DNA Topology*. New York: Oxford University Press.
5. **Batra D, Shea KJ.** (2003). Combinatorial methods in molecular imprinting. *Curr. Opin. Chem. Biol.* 7:434-42
6. **Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW.** (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D* 67:271–81
7. **Baugh C, Grate D, Wilson C.** (2000). 2.8 Å Crystal structure of the malachite green aptamer. *J. Mol. Biol.* 301:117-28
8. **Baugh L, Le Trong I, Cerutti DS, Glich S, Stayton PS, *et al.*** (2010). A distal point mutation in the streptavidin-biotin complex preserves structure but diminishes binding affinity: Experimental evidence of electronic polarization effects? *Biochemistry* 49:4568-70
9. **Beck A, Wurch T, Bailly C, Corvaia N.** (2010). Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* 10:345-52
10. **Beckett D, Kovaleva E, Schatz PJ.** (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 8:921-9
11. **Berg OG, Blomberg C.** (1976). Association kinetics with coupled diffusional flows. Special application to the lac repressor operator system. *Biophys. Chem.* 4:367-81
12. **Berg OG, Winter RB, von Hippel PH.** (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929-48
13. **Bermejo R, Lai M, Foiani M.** (2012). Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription. *Mol. Cell* 45:710-8
14. **Bing T, Yang X, Mei H, Cao Z, Shangguan D.** (2010). Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories. *Bioorg. Med. Chem.* 18:1798-805
15. **Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, *et al.*** (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* 22:575-82
16. **Binz HK, Amstutz P, Plückthun A.** (2005). Engineering novel binding proteins from non-immunoglobulin domains. *Nat. Biotechnol.* 23:1257-68
17. **Bloom L, Calabro V.** (2009). FN3: a new protein scaffold reaches the clinic. *Drug Discov. Today* 14:949–55
18. **Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ.** (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355:564-6
19. **Bondeson K, Frostell-Karlsson A, Fagerstam L, Magnusson G.** (1993). Lactose repressor-operator DNA interactions: Kinetic analysis by a surface plasmon resonance biosensor. *Anal. Biochem.* 214:245–51
20. **Branson BM.** (2004). FDA approves OraQuick for use in saliva. On March 25, the FDA approved the first rapid test for HIV in oral fluids. *AIDS Clin. Care* 16:39
21. **Bratkovič T.** (2010). Progress in phage display: Evolution of the technique and its applications. *Cell. Mol. Life Sci.* 67:749-67
22. **Bravman T, Bronner V, Nahshol O, Schreiber G.** (2008). The ProteOn XPR36™ Array System—High Throughput Kinetic Binding Analysis of Biomolecular Interactions. *Cell. Mol. Bioeng.* 1:216-28
23. **Browne SM, Al-Rubeai M.** (2007). Selection methods for high-producing mammalian cell lines. *Trends Biotechnol.* 25:425-32
24. **Bruno JG, Carrillo MP, Phillips T, King B.** (2008). Development of DNA aptamers for cytochemical detection of acetylcholine. *In Vitro Cell. Dev. Biol. Anim* 44:63-72
25. **Bugaut A, Toulmé JJ, Rayner B.** (2006). SELEX and dynamic combinatorial chemistry interplay for the selection of conjugated RNA aptamers. *Org. Biomol. Chem.* 4:4082-8
26. **Cabrera HA.** (1969). A comprehensive evaluation of pregnancy tests. *Am. J. Obstet. Gynecol.* 103:32-8
27. **Cao XX, Li SH, Chen LC, Ding HM, Xu H, *et al.*** (2009). Combining use of a panel of ssDNA aptamers in the detection of *Staphylococcus aureus*. *Nucleic Acids Res.* 37:4621-8
28. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, *et al.*** (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7:335-6
29. **Carter PJ.** (2006). Potent antibody therapeutics by design. *Nat. Rev. Immunol.* 6:343-57

30. **Cederfur J, Pei Y, Zihui M, Kempe M.** (2003). Synthesis and screening of a molecularly imprinted polymer library targeted for penicillin G. *J. Comb. Chem.* 5:67-72
31. **Chakravartty V, Cronan JE.** (2012). Altered regulation of *Escherichia coli* biotin biosynthesis in bira superrepressor mutant strains. *J. Bacteriol.* 194:1113-26
32. **Chambers RS, Johnston SA.** (2003). High-level generation of polyclonal antibodies by genetic immunization. *Nat. Biotechnol.* 21:1088-92
33. **Chapuis F, Pichon V, Lanza F, Sellergren S, Hennion MC.** (2003). Optimization of the class-selective extraction of triazines from aqueous samples using a molecularly imprinted polymer by a comprehensive approach of the retention mechanism. *J. Chromatogr.* 999:23-33
34. **Chayen NE, Saridakis E.** (2008). Protein crystallization: From purified protein to diffraction-quality crystal. *Nat. Methods* 5:147-53
35. **Chelliserrykattil J, Ellington AD.** (2004). Evolution of a T7 RNA polymerase variant that transcribes 2′-O-methyl RNA. *Nat. Biotechnol.* 22:1155-60
36. **Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, *et al.*** (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* 66:12-21
37. **Chiarella P, Fazio VM.** (2008). Mouse monoclonal antibodies in biological research: Strategies for high-throughput production. *Biotechnol. Lett.* 30:1303-10
38. **Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, *et al.*** (2010). Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* 107:15373
39. **Clark SL, Remcho VT.** (2002). Aptamers as analytical reagents. *Electrophoresis* 23:1335-40
40. **Convery MA, Rowsell S, Storehouse NJ, Ellington AD, Hirao I, *et al.*** (1998). Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* 5:133-9
41. **Cox JC, Hayhurst A, Hesselberth J, Bayer TS, Georgiou G, Ellington AD.** (2002). Automated selection of aptamers against protein targets translated *in vitro*: from gene to aptamer. *Nucleic Acids Res.* 30:e108
42. **Cruz-Aguado JA, Penner G.** (2008). Determination of ochratoxin A with a DNA aptamer. *J. Agric. Food Chem.* 56:10456-61
43. **Dang C, Jayasena SD.** (1996). Oligonucleotide inhibitors of Taq DNA polymerase facilitate detection of low copy number targets by PCR. *J. Mol. Biol.* 264:268-78
44. **Darling RJ, Brault PA.** (2004). Kinetic exclusion assay technology: Characterization of molecular interactions. *Assay Drug Dev. Technol.* 2:647-57
45. **Dassie JP, Liu XY, Thomas GS, Whitaker RM, Thiel KW, *et al.*** (2009). Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors. *Nat. Biotechnol.* 27:839-46
46. **Daugherty PS.** (2007). Protein engineering with bacterial display. *Curr. Opin. Struct. Biol.* 17:474-80
47. **Davies DR, Gelinas AD, Zhang C, Rohloff JC, Carter JD, *et al.*** (2012). Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. USA* 109:19971-6
48. **Davis JH, Szostak JW.** (2002). Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. USA* 99:11616-21
49. **Day YSN, Baird CL, Rich RL, Myszka DG.** (2002). Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. *Protein Sci.* 11:1017-25
50. **De Masi F, Chiarella P, Wilhelm H, Massimi M, Bullard B, *et al.*** (2005). High throughput production of mouse monoclonal antibodies using antigen microarrays. *Proteomics* 5:4070-81
51. **De Vos WM.** (2011). Systems solutions by lactic acid bacteria: From paradigms to practice. *Microb. Cell Fact.* 10(Suppl 1)
52. **Dickey FH.** (1949). The Preparation of Specific Adsorbents. *Proc. Natl. Acad. Sci. USA* 35:227-9
53. **Doudna JA, Grosshans C, Gooding A, Kundrot CE.** (1993). Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Sci. USA* 90:7829-33
54. **Drake AW, Myszka DG, Klakamp SL.** (2004). Characterizing high-affinity antigen/antibody complexes by kinetic- and equilibrium-based methods. *Anal. Biochem.* 328:35-43
55. **Drolet DW, Moon-McDermott L, Romig TS.** (1996). An enzyme-linked oligonucleotide assay. *Nat. Biotechnol.* 14:1021-7
56. **Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB.** (2008). The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* 36:5123-38
57. **Eismann ER, Müller-Hill B.** (1990). Lac repressor forms stable loops *in Vitro* with supercoiled wild-type lac DNA containing all three natural lac operators. *J. Mol. Biol.* 213:763-75
58. **Ellington AD, Szostak JW.** (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 346:818-22

59. **Escudé C, Garestier T, Hélène C.** (1999). Padlock oligonucleotides for duplex DNA based on sequence-specific triple helix formation. *Proc. Natl. Acad. Sci. USA* 96:10603-7
60. **Famulok M.** (1994). Molecular recognition of amino acids by RNA-aptamers: An L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* 116:1698-706
61. **FDA.** (2010). *FDA: Pfizer Voluntarily Withdraws Cancer Treatment Mylotarg from U.S. Market*. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm216448.htm
62. **Franco EJ, Sonneson GJ, DeLegge TJ, Hofstetter H, Horn JR, Hofstetter O.** (2010). Production and characterization of a genetically engineered anti-caffeine camelid antibody and its use in immunoaffinity chromatography. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 878:177-86
63. **Freyer MW, Lewis EA.** (2008). Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods Cell Biol.* 84:79-113
64. **Friedmann D, Messick T, Marmorstein R.** (2011). Crystallization of macromolecules. *Curr. Protoc. Protein Sci.* Chapter 17
65. **Fuller RS, Funnell BE, Kornberg A.** (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* 38:889-900
66. **Fuller RS, Kornberg A.** (1983). Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. USA* 80:5817-21
67. **Gai SA, Wittrup KD.** (2007). Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* 17:467-73
68. **Garber M, Gongadze G, Meshcheryakov V, Nikonov O, Nikulin A, *et al.*** (2002). Crystallization of RNA/protein complexes. *Acta Crystallogr. Sect. D* 58:1664-9
69. **Ge Y, Turner APF.** (2008). Too large to fit? Recent developments in macromolecular imprinting. *Trends Biotechnol.* 26:218-24
70. **Ge Y, Turner APF.** (2009). Molecularly imprinted sorbent assays: Recent developments and applications. *Chem. Eur. J.* 15:8100-7
71. **Geiger A, Burgstaller P, Von der Eltz H, Roeder A, Famulok M.** (1996). RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.* 24:1029-36
72. **Géron-Landre B, Roulon T, Escudé C.** (2005). Stem-loop oligonucleotides as tools for labelling double-stranded DNA. *FEBS J.* 272:5343-52
73. **Ghai R, Falconer RJ, Collins BM.** (2012). Applications of isothermal titration calorimetry in pure and applied research-survey of the literature from 2010. *J. Mol. Recognit.* 25:32-52
74. **Gill DS, Damle NK.** (2006). Biopharmaceutical drug discovery using novel protein scaffolds. *Curr. Opin. Biotechnol.* 17:653-8
75. **Girolamo AD, Le L, Penner G, Schena R, Visconti A.** (2012). Analytical performances of a DNA-ligand system using time-resolved fluorescence for the determination of ochratoxin A in wheat. *Anal. Bioanal. Chem.* 403:2627-34
76. **Goeddel DV, Yansura DG, Caruthers MH.** (1977). Binding of synthetic lactose operator DNAs to lactose repressors. *Proc. Natl. Acad. Sci. USA* 74:3292-6
77. **Gold L, Ayers D, Bertino J, Bock C, Bock A, *et al.*** (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* 5
78. **Gold L, Janjic N, Jarvis T, Schneider D, Walker JJ, *et al.*** (2012). Aptamers and the RNA World, Past and Present. *Cold Spring Harb. Perspect. Biol.* 4
79. **Goldstein G, Schindler J, Tsai H.** (1985). A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. *New Engl. J. Med.* 313:337-42
80. **Gomes J, Steiner W.** (2004). The biocatalytic potential of extremophiles and extremozymes. *Food Technol. Biotechnol.* 42:223-35
81. **Gowers DM, Fox KR.** (1999). Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.* 27:1569-77
82. **Gragoudas ES, Adamis AP, Cunningham Jr ET, Feinsod M, Guyer DR.** (2004). Pegaptanib for neovascular age-related macular degeneration. *New Engl. J. Med.* 351:2805-16
83. **Grönwall C, Ståhl S.** (2009). Engineered affinity proteins-Generation and applications. *J. Biotechnol.* 140:254-69
84. **Halford SE, Gowers DM, Sessions RB.** (2000). Two are better than one. *Nat. Struct. Biol.* 7:705-7
85. **Halford SE, Marko JF.** (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040-52
86. **Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, *et al.*** (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363:446-8
87. **Harmsen MM, De Haard HJ.** (2007). Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* 77:13-22

88. **Hendrickx APA, Budzik JM, Oh SY, Schneewind O.** (2011). Architects at the bacterial surface-sortases and the assembly of pili with isopeptide bonds. *Nat. Rev. Microbiol.* 9:166-76
89. **Hermann T, Patel DJ.** (2000). Adaptive recognition by nucleic acid aptamers. *Science* 287:820-5
90. **Hey T, Fiedler E, Rudolph R, Fiedler M.** (2005). Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends Biotechnol.* 23:514-22
91. **Hianik T, Ostatná V, Sonlajtnerova M, Grman I.** (2007). Influence of ionic strength, pH and aptamer configuration for binding affinity to thrombin. *Bioelectrochemistry* 70:127-33
92. **Hober S, Nord K, Linhult M.** (2007). Protein A chromatography for antibody purification. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 848:40-7
93. **Hoggan DB, Chao JA, Prasad GS, Stout CD, Williamson JR.** (2003). Combinatorial crystallization of an RNA-protein complex. *Acta Crystallogr. Sect. D* 59:466-73
94. **Hoinka J, Zotenko E, Friedman A, Sauna ZE, Przytycka TM.** (2012). Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics* 28:i215-i23
95. **Holliger P, Hudson PJ.** (2005). Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* 23:1126-36
96. **Hollis T.** (2007). Crystallization of protein-DNA complexes. *Methods Mol. Biol.* 363:225-37
97. **Holthausen JT, Wyman C, Kanaar R.** (2010). Regulation of DNA strand exchange in homologous recombination. *DNA Repair* 9:1264-72
98. **Holthoff EL, Bright FV.** (2007). Molecularly templated materials in chemical sensing. *Anal. Chim. Acta* 594:147-61
99. **Homola J.** (2006). Electromagnetic Theory of Surface Plasmons. In *Surface Plasmon Resonance Based Sensors*, ed. J Homola, 4:3-44. Heidelberg: Springer Berlin Heidelberg.
100. **Homola J, Piliarik M.** (2006). Surface Plasmon Resonance (SPR) Sensors. In *Surface Plasmon Resonance Based Sensors*, ed. J Homola, 4:45-67: Springer Berlin Heidelberg.
101. **Hoogenboom HR.** (2005). Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* 23:1105-16
102. **Hoon S, Zhou B, Janda KD, Brenner S, Scolnick J.** (2011). Aptamer selection by high-throughput sequencing and informatic analysis. *BioTechniques* 51:413-6
103. **Horn WT, Convery MA, Stonehouse NJ, Adams CJ, Liljas L, *et al.*** (2004). The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: Further challenges in the modeling of ligand-RNA interactions. *RNA* 10:1776-82
104. **Hoshino Y, Koide H, Urakami T, Kanazawa H, Kodama T, *et al.*** (2010). Recognition, Neutralization, and Clearance of Target Peptides in the Bloodstream of Living Mice by Molecularly Imprinted Polymer Nanoparticles: A Plastic Antibody. *J. Am. Chem. Soc.* 132:6644-5
105. **Hosse RJ, Rothe A, Power BE.** (2006). A new generation of protein display scaffolds for molecular recognition. *Protein Sci.* 15:14-27
106. **Hu X, Hu H, Melvin JA, Clancy KW, McCafferty DG, Yang W.** (2011). Autocatalytic intramolecular isopeptide bond formation in Gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* 133:478-85
107. **Huang DB, Vu D, Cassiday LA, Zimmerman JM, Maher III LJ, Ghosh G.** (2003). Crystal structure of NF-κb (p50)2 complexed to a high-affinity RNA aptamer. *Proc. Natl. Acad. Sci. USA* 100:9268-73
108. **Huang RH, Fremont DH, Diener JL, Schaub RG, Sadler JE.** (2009). A Structural Explanation for the Antithrombotic Activity of ARC1172, a DNA Aptamer that Binds von Willebrand Factor Domain A1. *Structure* 17:1476-84
109. **Huber W, Mueller F.** (2006). Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. *Curr. Pharm. Des.* 12:3999-4021
110. **Huizenga DE, Szostak JW.** (1995). A DNA aptamer that binds adenosine and ATP. *Biochemistry* 34:656-65
111. **Hura GL, Menon AL, Hammel M, Rambo RP, Poole II FL, *et al.*** (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* 6:606-12
112. **Huval CC, Bailey MJ, Braunlin WH, Randall Holmes-Farley S, Harry Mandeville W, *et al.*** (2001). Novel cholesterol lowering polymeric drugs obtained by molecular imprinting. *Macromolecules* 34:1548-50
113. **Hwang J, Fauzi H, Fukuda K, Sekiya S, Kakiuchi N, *et al.*** (2000). The RNA aptamer-binding site of hepatitis C virus NS3 protease. *Biochem. Biophys. Res. Commun.* 279:557-62
114. **Jackson GW.** (2012). *Nucleic Acid Aptamers for Diagnostics and Therapeutics: Global Markets*. BCC Research.
115. **Jäger S, Rasched G, Kornreich-Leshem H, Engeser M, Thum O, Famulok M.** (2005). A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.* 127:15071-82
116. **Janeway CAJ, Travers P, Walport M.** (2001). The Immune System in Health and Disease. In *Immunobiology*, 5. New York: Garland Science.

117. **Javaherian S, Musheev MU, Kanoatov M, Berezovski MV, Krylov SN.** (2009). Selection of aptamers for a protein target in cell lysate and their application to protein purification. *Nucleic Acids Res.* 37:e62
118. **Jenison RD, Gill SC, Pardi A, Polisky B.** (1994). High-resolution molecular discrimination by RNA. *Science* 263:1425-9
119. **Jerabek-Willemsen M, Wienken CJ, Braun D, Baaske P, Duhr S.** (2011). Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* 9:342-53
120. **Jhaveri S, Rajendran M, Ellington AD.** (2000). *In vitro* selection of signaling aptamers. *Nat. Biotechnol.* 18:1293-7
121. **Jiang X, Egli M.** (2011). Use of chromophoric ligands to visually screen co-crystals of putative protein-nucleic acid complexes. *Curr. Protoc. Nucleic Acid Chem.*
122. **Joss L, Morton TA, Doyle ML, Myszka DG.** (1998). Interpreting kinetic rate constants from optical biosensor data recorded on a decaying surface. *Anal. Biochem.* 261:203-10
123. **Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN.** (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* 318:1625-8
124. **Kankainen M, Paulin L, Tynkkynen S, Von Ossowski I, Reunanen J, *et al.*** (2009). Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human-mucus binding protein. *Proc. Natl. Acad. Sci. USA* 106:17193-8
125. **Karim K, Breton F, Rouillon R, Piletska EV, Guerreiro A, *et al.*** (2005). How to find effective functional monomers for effective molecularly imprinted polymers? *Adv. Drug Del. Rev.* 57:1795-808
126. **Karlsson R, Katsamba PS, Nordin H, Pol E, Myszka DG.** (2006). Analyzing a kinetic titration series using affinity biosensors. *Anal. Biochem.* 349:136-47
127. **Kato T, Takemura T, Yano K, Ikebukuro K, Karube I.** (2000). *In vitro* selection of DNA aptamers which bind to cholic acid. *Biochim. Biophys. Acta* 1493:12-8
128. **Kaur H, Yung L-YL.** (2012). Probing High Affinity Sequences of DNA Aptamer against VEGF 165. *PLoS ONE* 7
129. **Ke A, Doudna JA.** (2004). Crystallization of RNA and RNA-protein complexes. *Methods* 34:408-14
130. **Keefe AD, Cload ST.** (2008). SELEX with modified nucleotides. *Curr. Opin. Chem. Biol.* 12:448-56
131. **Keefe AD, Pai S, Ellington A.** (2010). Aptamers as therapeutics. *Nat. Rev. Drug Discov.* 9:537-50
132. **Keeney TR, Bock C, Gold L, Kraemer S, Lollo B, *et al.*** (2009). Automation of the SomaLogic Proteomics Assay: A Platform for Biomarker Discovery. *JALA* 14:360-6
133. **Kehoe JW, Kay BK.** (2005). Filamentous phage display in the new millennium. *Chem. Rev.* 105:4056–72
134. **Kelly JA, Feigon J, Yeates TO.** (1996). Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.* 256:417-22
135. **Kikin O, D'Antonio L, Bagga PS.** (2006). QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34:W676-W82
136. **Kim YS, Jung HS, Matsuura T, Lee HY, Kawai T, Gu MB.** (2007). Electrochemical detection of 17 beta-estradiol using DNA aptamer immobilized gold electrode chip. *Biosensors Bioelectron.* 22:2525–31
137. **Klussmann S, Nolte A, Bald R, Erdmann VA, Fürste JP.** (1996). Mirror-image RNA that binds D-adenosine. *Nat. Biotechnol.* 14:1112-5
138. **Knight R, Yarus M.** (2003). Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res.* 31:e30
139. **Kohler G, Milstein C.** (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256:495-7
140. **Koide A, Bailey CW, Huang X, Koide S.** (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* 284:1141-51
141. **Konarev PV, Petoukhov MV, Volkov VV, Svergun DI.** (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* 39:277-86
142. **Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI.** (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36:1277-82
143. **Kontermann R, Dübel S, Storz U.** (2010). IP Issues in the Therapeutic Antibody Industry. In *Antibody Engineering*:517-81. Berlin: Springer.
144. **Krauss IR, Merlino A, Giancola C, Randazzo A, Mazzarella L, Sica F.** (2011). Thrombin-aptamer recognition: A revealed ambiguity. *Nucleic Acids Res.* 39:7858-67
145. **Kuwahara M, Hanawa K, Ohsawa K, Kitagata R, Ozaki H, Sawai H.** (2006). Direct PCR amplification of various modified DNAs having amino acids: Convenient preparation of DNA libraries with high-potential activities for *in vitro* selection. *Bioorg. Med. Chem.* 14:2518-26
146. **Kwan AH, Mobli M, Gooley PR, King GF, MacKay JP.** (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* 278:687-703
147. **Laing C, Schlick T.** (2011). Computational approaches to RNA structure prediction, analysis, and design. *Curr. Opin. Struct. Biol.* 21:306-18

148. **Laitinen OH, Hytönen VP, Nordlund HR, Kulomaa MS.** (2006). Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* 63:2992-3017
149. **Latham JA, Johnson R, Toole JJ.** (1994). The application of a modified nucleotide in aptamer selection: Novel thrombin aptamers containing 5-(1-pentynyl)-2′-deoxyuridine. *Nucleic Acids Res.* 22:2817-22
150. **Lauhon CT, Szostak JW.** (1995). RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.* 117:1246-57
151. **Le Trong I, Wang Z, Hyre DE, Lybrand TP, Stayton PS, Stenkamp RE.** (2011). Streptavidin and its biotin complex at atomic resolution. *Acta Crystallogr. Sect. D* 67:813-21
152. **Lebruska LL, Maher III LJ.** (1999). Selection and characterization of an RNA decoy for transcription factor NF-κB. *Biochemistry* 38:3168-74
153. **Lee YJ, Han SR, Maeng JS, Cho YJ, Lee SW.** (2012). *In vitro* selection of *Escherichia coli* O157:H7-specific RNA aptamer. *Biochem. Biophys. Res. Commun.* 417:414-20
154. **Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, *et al.*** (1996). Crystal structure of the lactose operon represser and its complexes with DNA and inducer. *Science* 271:1247-54
155. **Li D, Song S, Fan C.** (2010). Target-responsive structural switching for nucleic acid-based sensors. *Acc. Chem. Res.* 43:631-41
156. **Lipovsek D, Plückthun A.** (2004). In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* 290:51-67
157. **Lonberg N.** (2008). Fully human antibodies from transgenic mouse and phage display platforms. *Curr. Opin. Immunol.* 20:450-9
158. **Long SB, Long MB, White RR, Sullenger BA.** (2008). Crystal structure of an RNA aptamer bound to thrombin. *RNA* 14:2504-12
159. **Lozupone CA, Knight R.** (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol. Rev.* 32:557-78
160. **Lupold SE, Hicke BJ, Lin Y, Coffey DS.** (2002). Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Res.* 62:4029-33
161. **Lutterotti A, Martin R.** (2008). Getting specific: monoclonal antibodies in multiple sclerosis. *Lancet Neurol.* 7:538-47
162. **Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J.** (1993). Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. USA* 90:3745-9
163. **Mairal T, Cengiz Özalp V, Lozano Sánchez P, Mir M, Katakis I, O'Sullivan CK.** (2008). Aptamers: Molecular tools for analytical applications. *Anal. Bioanal. Chem.* 390:989-1007
164. **Majka J, Speck C.** (2007). Analysis of protein-DNA interactions using surface plasmon resonance. *Adv. Biochem. Eng. Biot.* 104:13-36
165. **Mann D, Reinemann C, Stoltenburg R, Strehlitz B.** (2005). *In vitro* selection of DNA aptamers binding ethanolamine. *Biochem. Biophys. Res. Commun.* 338:1928-34
166. **Martí AA, Jockusch S, Li Z, Ju J, Turro NJ.** (2006). Molecular beacons with intrinsically fluorescent nucleotides. *Nucleic Acids Res.* 34:e50
167. **Mascini M, Palchetti I, Tombelli S.** (2012). Nucleic acid and peptide aptamers: Fundamentals and bioanalytical aspects. *Angew. Chem. Int. Ed.* 51:1316-32
168. **Mayer G.** (2009). The chemical biology of aptamers. *Angew. Chem. Int. Ed.* 48:2672-89
169. **Mayes AG, Whitcombe MJ.** (2005). Synthetic strategies for the generation of molecularly imprinted organic polymers. *Adv. Drug Del. Rev.* 57:1742-78
170. **McCauley TG, Hamaguchi N, Stanton M.** (2003). Aptamer-based biosensor arrays for detection and quantification of biological macromolecules. *Anal. Biochem.* 319:244-50
171. **McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.** (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40:658-74
172. **McNamara II JO, Andrechek ER, Wang Y, Viles KD, Rempel RE, *et al.*** (2006). Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. *Nat. Biotechnol.* 24:1005-15
173. **McPherson A.** (2004). Introduction to protein crystallization. *Methods* 34:254-65
174. **Meares CF.** (2009). Targeted instant immunity. *Nat. Biotechnol.* 27:452-3
175. **Mehta J, Rouah-Martin E, Van Dorst B, Maes B, Herrebout W, *et al.*** (2011). Selection and Characterization of PCB-Binding DNA Aptamers. *Anal. Chem.* 84:1669–76
176. **Mehta J, Van Dorst B, Rouah-Martin E, Herrebout W, Scippo ML, *et al.*** (2011). *In vitro* selection and characterization of DNA aptamers recognizing chloramphenicol. *J. Biotechnol.* 155:361-9
177. **Mendonsa SD, Bowser MT.** (2004). *In vitro* selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. *Anal. Chem.* 76:5387-92
178. **Metzker ML.** (2010). Sequencing technologies the next generation. *Nat. Rev. Genet.* 11:31-46
179. **Missailidis S, Hardy A.** (2009). Aptamers as inhibitors of target proteins. *Expert Opin. Ther. Pat.* 19:1073-82

180. **Miyakawa S, Nomura Y, Sakamoto T, Yamaguchi Y, Kato K, *et al.*** (2008). Structural and molecular basis for hyperspecificity of RNA aptamer to human immunoglobulin G. *RNA* 14:1154-63
181. **Montiel C, Bustos-Jaimes I.** (2008). Trends and challenges in directed evolution. *Curr. Chem. Biol.* 2:50-9
182. **Moorthy AK, Huang DB, Wang VYF, Vu D, Ghosh G.** (2007). X-ray Structure of a NF-κB p50/RelB/DNA Complex Reveals Assembly of Multiple Dimers on Tandem κB Sites. *J. Mol. Biol.* 373:723-34
183. **Morton TA, Myszka DG.** (1998). Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Methods Enzymol.* 295:268-94
184. **Mossing MC, Record Jr MT.** (1986). Upstream operators enhance repression of the lac promoter. *Science* 233:889-92
185. **Müller KM, Arndt KM, Plückthun A.** (1998). Model and simulation of multivalent binding to fixed ligands. *Anal. Biochem.* 261:149-58
186. **Mullis KB.** (2008). Chemically programmable immunity. *Patent No. US 7,422,746*
187. **Murai R, Yoshikawa HY, Kawahara H, Maki S, Sugiyama S, *et al.*** (2008). Effect of solution flow produced by rotary shaker on protein crystallization. *J. Cryst. Growth* 310:2168-72
188. **Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, *et al.*** (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D* 67:355-67
189. **Myszka DG.** (1999). Improving biosensor analysis. *J. Mol. Recognit.* 12:279-84
190. **Myszka DG.** (1999). Survey of the 1998 optical biosensor literature. *J. Mol. Recognit.* 12:390-408
191. **Myszka DG.** (2000). Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Methods Enzymol.* 323:325-40
192. **Myszka DG, He X, Dembo M, Morton TA, Goldstein B.** (1998). Extending the range of rate constants available from BIACORE: Interpreting mass transport-influenced binding data. *Biophys. J.* 75:583-94
193. **Myszka DG, Morton TA.** (1998). Clamp: A biosensor kinetic data analysis program. *Trends Biochem. Sci.* 23:149-50
194. **Neidle S.** (2009). The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* 19:239-50
195. **Nelson AL, Dhimolea E, Reichert JM.** (2010). Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* 9:767-74
196. **Ng EWM, Shima DT, Calias P, Cunningham Jr ET, Guyer DR, Adamis AP.** (2006). Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discov.* 5:123-32
197. **Niazi JH, Lee SJ, Gu MB.** (2008). Single-stranded DNA aptamers specific for antibiotics tetracyclines. *Bioorg. Med. Chem.* 16:7245-53
198. **Ning Y, Wang Y, Li Y, Hong Y, Peng D, *et al.*** (2006). An alternative strategy for high throughput generation and characterization of monoclonal antibodies against human plasma proteins using fractionated native proteins as immunogens. *Proteomics* 6:438-48
199. **Nix J, Sussman D, Wilson C.** (2000). The 1.3 Å crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. *J. Mol. Biol.* 296:1235-44
200. **Nomura Y, Sugiyama S, Sakamoto T, Miyakawa S, Adachi H, *et al.*** (2010). Conformational plasticity of RNA for target recognition as revealed by the 2.15 Å crystal structure of a human IgG-aptamer complex. *Nucleic Acids Res.* 38:7822-9
201. **Normanno D, Vanzi F, Pavone FS.** (2008). Single-molecule manipulation reveals supercoiling-dependent modulation of lac repressor-mediated DNA looping. *Nucleic Acids Res.* 36:2505-13
202. **Nutiu R, Li Y.** (2005). *In vitro* selection of structure-switching signaling aptamers. *Angew. Chem. Int. Ed.* 44:1061-5
203. **Nuttall SD, Walsh RB.** (2008). Display scaffolds: protein engineering for novel therapeutics. *Curr. Opin. Pharm.* 8:609-15
204. **Nygren PÅ, Skerra A.** (2004). Binding proteins from alternative scaffolds. *J. Immunol. Methods* 290:3-28
205. **Oehler S, Eismann ER, Kramer H, Muller-Hill B.** (1990). The three operators of the lac operon cooperate in repression. *EMBO J.* 9:973-9
206. **Oh SS, Plakos K, Lou X, Xiao Y, Soh HT.** (2010). *In vitro* selection of structure-switching, self-reporting aptamers. *Proc. Natl. Acad. Sci. USA* 107:14053-8
207. **Oktem HA, Bayramoglu G, Ozalp VC, Arica MY.** (2007). Single-step purification of recombinant Thermus aquaticus DNA polymerase using DNA-aptamer immobilized novel affinity magnetic beads. *Biotechnol. Prog.* 23:146-54
208. **Orlova EV, Saibil HR.** (2011). Structural analysis of macromolecular assemblies by electron microscopy. *Chem. Rev.* 111:7710-48
209. **Padilla R, Sousa R.** (2002). A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.* 30:e138

210. **Padmanabhan K, Padmanabhan KP, Ferrara JD, Sadler JE, Tulinsky A.** (1993). The structure of α-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* 268:17651-4
211. **Padmanabhan K, Tulinsky A.** (1996). An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Crystallogr. Sect. D* 52:272-82
212. **Paige JS, Wu KY, Jaffrey SR.** (2011). RNA Mimics of Green Fluorescent Protein. *Science* 333:642-6
213. **Pande J, Szewczyk MM, Grover AK.** (2010). Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* 28:849-58
214. **Papalia G, Myszka D.** (2010). Exploring minimal biotinylation conditions for biosensor analysis using capture chips. *Anal. Biochem.* 403:30-5
215. **Parisien M, Major F.** (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51-5
216. **Park SJ, Cochran JR.** (2010). *Protein engineering and design*. Boca Raton: CRC Press.
217. **Paschke M.** (2006). Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* 70:2-11
218. **Patek M, Drew M.** (2008). Chemical synthesis in nanosized cavities. *Curr. Opin. Chem. Biol.* 12:332-9
219. **Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, *et al.*** (2012). Synthetic genetic polymers capable of heredity and evolution. *Science* 336:341-4
220. **Poniková S, Tlučková K, Antalík M, Víglaský V, Hianik T.** (2011). The circular dichroism and differential scanning calorimetry study of the properties of DNA aptamer dimers. *Biophys. Chem.* 155:29-35
221. **Popkov M, Gonzalez B, Sinha SC, Barbas III CF.** (2009). Instant immunity through chemically programmable vaccination and covalent self-assembly. *Proc. Natl. Acad. Sci. USA* 106:4378-83
222. **Ramette A.** (2007). Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* 62:142-60
223. **Rathbone DL.** (2005). Molecularly imprinted polymers in the drug discovery process. *Adv. Drug Del. Rev.* 57:1854-74
224. **Reichert JM.** (2008). Monoclonal antibodies as innovative therapeutics. *Curr. Pharm. Biotechnol.* 9:423-30
225. **Reinemann C, Stoltenburg R, Strehlitz B.** (2009). Investigations on the Specificity of DNA Aptamers Binding to Ethanolamine. *Anal. Chem.* 81:3973-8
226. **Reinstein O, Neves MAD, Saad M, Boodram SN, Lombardo S, *et al.*** (2011). Engineering a structure switching mechanism into a steroid-binding aptamer and hydrodynamic analysis of the ligand binding mechanism. *Biochemistry* 50:9368-76
227. **Renault L, Chou HT, Chiu PL, Hill RM, Zeng X, *et al.*** (2006). Milestones in electron crystallography. *J. Comput.-Aided Mol. Des.* 20:519-27
228. **Reunanen J, von Ossowski I, Hendrickx APA, Palva A, de Vosa WM.** (2012). Characterization of the SpaCBA pilus fibers in the probiotic Lactobacillus rhamnosus GG. *Appl. Environ. Microbiol.* 78:2337-44
229. **Rich RL, Myszka DG.** (2007). Survey of the year 2006 commercial optical biosensor literature. *J. Mol. Recognit.* 20:300-66
230. **Rich RL, Myszka DG.** (2008). Survey of the year 2007 commercial optical biosensor literature. *J. Mol. Recognit.* 21:355-400
231. **Rich RL, Myszka DG.** (2009). Extracting affinity constants from biosensor binding responses. In *Label-free Biosensors: Techniques and Applications*, ed. MA Cooper:48-84. Cambridge: Cambridge University Press.
232. **Rich RL, Myszka DG.** (2009). Extracting kinetic rate constants from biosensor binding responses. In *Label-free Biosensors: Techniques and Applications*, ed. MA Cooper:85-109. Cambridge: Cambridge University Press.
233. **Rich RL, Myszka DG.** (2010). Grading the commercial optical biosensor literature - Class of 2008: 'The Mighty Binders'. *J. Mol. Recognit.* 23:1-64
234. **Rich RL, Myszka DG.** (2011). The Revolution of Real-Time, Label-Free Biosensor Applications. In *Label-Free Technologies for Drug Discovery*, ed. MA Cooper, LM Mayr:1-25. Chichester: John Wiley & Sons, Ltd.
235. **Rich RL, Myszka DG.** (2011). Survey of the 2009 commercial optical biosensor literature. *J. Mol. Recognit.* 24:892-914
236. **Rich RL, Papalia GA, Flynn PJ, Furneisen J, Quinn J, *et al.*** (2009). A global benchmark study using affinity-based biosensors. *Anal. Biochem.* 386:194-216
237. **Richter PH, Eigen M.** (1974). Diffusion controlled reaction rates in spheroidal geometry. Application to repressor: operator association and membrane repressor enzymes. *Biophys. Chem.* 2:255-63
238. **Riggs AD, Bourgeois S, Cohn M.** (1970). The lac represser-operator interaction. III. Kinetic studies. *J. Mol. Biol.* 53:401-17
239. **Ritzefeld M, Sewald N.** (2012). Real-Time Analysis of Specific Protein-DNA Interactions with Surface Plasmon Resonance. *J. amino acids* 2012:1-19
240. **Rivas E, Eddy SR.** (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* 285:2053-68

241. **Romig TS, Bell C, Drolet DW.** (1999). Aptamer affinity chromatography: Combinatorial chemistry applied to protein purification. *J. Chromatogr. B Biomed. Sci. Appl.* 731:275-84
242. **Roulon T, Coulaud D, Delain E, Le Cam E, Hélène C, Escudé C.** (2002). Padlock oligonucleotides as a tool for labeling superhelical DNA. *Nucleic Acids Res.* 30:e12
243. **Rowsell S, Stonehouse NJ, Convery MA, Adams CJ, Ellington AD, *et al.*** (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat. Struct. Biol.* 5:970-5
244. **Ruckman J, Green LS, Beeson J, Waugh S, Gillette WL, *et al.*** (1998). 2'-fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF 165): Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* 273:20556-67
245. **Ruigrok VJB, Levisson M, Eppink MHM, Smidt H, van der Oost J.** (2011). Alternative affinity tools: more attractive than antibodies? *Biochem. J.* 436:1-13
246. **Ruigrok VJB, Levisson M, Hekelaar J, Smidt H, Dijkstra BW, van der Oost J.** (2012). Characterization of Aptamer-Protein Complexes by X-ray Crystallography and Alternative Approaches. *Int. J. Mol. Sci.* 13:10537-52
247. **Ruigrok VJB, van Duijn E, Barendregt A, Dyer K, Tainer JA, *et al.*** (2012). Kinetic and Stoichiometric Characterisation of Streptavidin-Binding Aptamers. *ChemBioChem* 13:829-36
248. **Ruta J, Ravelet C, Désiré J, Décout JL, Peyrin E.** (2008). Covalently bonded DNA aptamer chiral stationary phase for the chromatographic resolution of adenosine. *Anal. Bioanal. Chem.* 390:1051-7
249. **Ruusala T, Crothers DM.** (1992). Sliding and intermolecular transfer of the lac repressor: Kinetic perturbation of a reaction intermediate by a distant DNA sequence. *Proc. Natl. Acad. Sci. USA* 89:4903-7
250. **Sassanfar M, Szostak JW.** (1993). An RNA motif that binds ATP. *Nature* 364:550-3
251. **Saxelin M, Tynkkynen S, Mattila-Sandholm T, De Vos WM.** (2005). Probiotic and other functional microbes: From markets to mechanisms. *Curr. Opin. Biotechnol.* 16:204-11
252. **Schneidman-Duhovny D, Hammel M, Sali A.** (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 38:W540-W4
253. **Schütze T, Wilhelm B, Greiner N, Braun H, Peter F, *et al.*** (2011). Probing the SELEX process with next-generation sequencing. *PLoS ONE* 6
254. **Sekiya S, Noda K, Nishikawa F, Yokoyama T, Kumar PKR, Nishikawa S.** (2006). Characterization and application of a novel RNA aptamer against the mouse prion protein. *J. Biochem.* 139:383-90
255. **Shendure J, Aiden EL.** (2012). The expanding scope of DNA sequencing. *Nat. Biotechnol.* 30:1084-94
256. **Shi H, Tsal WB, Garrison MD, Ferrari S, Ratner BD.** (1999). Template-imprinted nanostructured surfaces for protein recognition. *Nature* 398:593-7
257. **Shum KT, Lui ELH, Wong SCK, Yeung P, Sam L, *et al.*** (2011). Aptamer-mediated inhibition of mycobacterium tuberculosis polyphosphate kinase 2. *Biochemistry* 50:3261-71
258. **Sidhu SS.** (2001). Engineering M13 for phage display. *Biomol. Eng.* 18:57-63
259. **Sidhu SS, Koide S.** (2007). Phage display for engineering and analyzing protein interaction interfaces. *Curr. Opin. Struct. Biol.* 17:481-7
260. **Skerra A.** (2007). Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol.* 18:295-304
261. **Skrzypczak-Jankun E, Carperos VE, Ravichandran KG, Tulinsky A, Westbrook M, Maraganore JM.** (1991). Structure of the hirugen and hirulog 1 complexes of α-thrombin. *J. Mol. Biol.* 221:1379-93
262. **Snyder DA, Chen Y, Denissova NG, Acton T, Aramini JM, *et al.*** (2005). Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *J. Am. Chem. Soc.* 127:16505-11
263. **Someya T, Baba S, Fujimoto M, Kawai G, Kumasaka T, Nakamura K.** (2011). Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res.*
264. **Song KM, Cho M, Jo H, Min K, Jeon SH, *et al.*** (2011). Gold nanoparticle-based colorimetric detection of kanamycin using a DNA aptamer. *Anal. Biochem.* 415:175-81
265. **Spivak DA.** (2005). Optimization, evaluation, and characterization of molecularly imprinted polymers. *Adv. Drug Del. Rev.* 57:1779-94
266. **Stanfield RL, Dooley H, Flajnik MF, Wilson IA.** (2004). Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* 305:1770-3
267. **Steck TR, Franco RJ, Wang JY, Drlica K.** (1993). Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. *Mol. Microbiol.* 10:473-81
268. **Stoltenburg R, Reinemann C, Strehlitz B.** (2005). FluMag-SELEX as an advantageous method for DNA aptamer selection. *Anal. Bioanal. Chem.* 383:83-91
269. **Stoltenburg R, Reinemann C, Strehlitz B.** (2007). SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* 24:381-403

270. **Strehlitz B, Reinemann C, Linkorn S, Stoltenburg R.** (2012). Aptamers for pharmaceuticals and their application in environmental analytics. *Bioanal. Rev.* 4:1-30
271. **Stumpp MT, Amstutz P.** (2007). DARPins: A true alternative to antibodies. *Curr. Opin. Drug Discovery Dev.* 10:153-9
272. **Stumpp MT, Binz HK, Amstutz P.** (2008). DARPins: A new generation of protein therapeutics. *Drug Discov. Today* 13:695-701
273. **Sugiyama S, Nomura Y, Sakamoto T, Kitatani T, Kobayashi A, *et al.*** (2008). Crystallization and preliminary X-ray diffraction studies of an RNA aptamer in complex with the human IgG Fc fragment. *Acta Crystallogr. Sect. F* 64:942-4
274. **Sun BW, Babu BR, Sørensen MD, Zakrzewska K, Wengel J, Sun JS.** (2004). Sequence and pH Effects of LNA-Containing Triple Helix-Forming Oligonucleotides: Physical Chemistry, Biochemistry, and Modeling Studies. *Biochemistry* 43:4160-9
275. **Sundaram P, Kurniawan H, Byrne ME, Wower J.** (2012). Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci.* 48:259-71
276. **Sussman D, Nix J, Wilson C.** (2000). The structural basis for molecular recognition by the vitamin B12 RNA aptamer. *Nat. Struct. Biol.* 7:53-7
277. **Tafvizi A, Mirny LA, Van Oijen AM.** (2011). Dancing on DNA: Kinetic aspects of search processes on DNA. *ChemPhysChem* 12:1481-9
278. **Tahallah N, Pinkse M, Maier CS, Heck AJR.** (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.* 15:596-601
279. **Takeuchi T, Fukuma D, Matsui J.** (1999). Combinatorial molecular imprinting: An approach to synthetic polymer receptors. *Anal. Chem.* 71:285-90
280. **Takeuchi T, Hishiya T.** (2008). Molecular imprinting of proteins emerging as a tool for protein recognition. *Org. Biomol. Chem.* 6:2459-67
281. **Tang ZW, Shangguan D, Wang KM, Shi H, Sefah K, *et al.*** (2007). Selection of aptamers for molecular recognition and characterization of cancer cells. *Anal. Chem.* 79:4900-7
282. **Taussig MJ, Stoevesandt O, Borrebaeck CAK, Bradbury AR, Cahill D, *et al.*** (2007). ProteomeBinders: Planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods* 4:13-7
283. **Tereshko V, Skripkin E, Patel DJ.** (2003). Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol.* 10:175-87
284. **Thiel K.** (2004). Oligo oligarchy - The surprisingly small world of aptamers. *Nat. Biotechnol.* 22:649-51
285. **Tuerk C, Gold L.** (1990). Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505-10
286. **Tuerk C, MacDougal S, Gold L.** (1992). RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* 89:6988-92
287. **Turiel E, Martín-Esteban A.** (2010). Molecularly imprinted polymers for sample preparation: A review. *Anal. Chim. Acta* 668:87-99
288. **Tyagi S, Kramer FR.** (1996). Molecular beacons: Probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14:303-8
289. **Unsworth LD, Van Der Oost J, Koutsopoulos S.** (2007). Hyperthermophilic enzymes - Stability, activity and implementation strategies for high temperature applications. *FEBS J.* 274:4044-56
290. **Vaught JD, Bock C, Carter J, Fitzwater T, Otis M, *et al.*** (2010). Expanding the chemistry of DNA for *in vitro* selection. *J. Am. Chem. Soc.* 132:4141-51
291. **Vlatakis G, Andersson LI, Muller R, Mosbach K.** (1993). Drug assay using antibody mimics made by molecular imprinting. *Nature* 361:645-7
292. **Vogel M, Keller-Gautschi E, Baumann MJ, Amstutz P, Ruf C, *et al.*** (2007). Designed ankyrin repeat proteins as anti-idiotypic-binding molecules. *Ann. N. Y. Acad. Sci.* 1109:9-18
293. **Von Lode P.** (2005). Point-of-care immunotesting: Approaching the analytical performance of central laboratory methods. *Clin. Biochem.* 38:591-606
294. **Von Ossowski I, Reunanen J, Satokari R, Vesterlund S, Kankainen M, *et al.*** (2010). Mucosal adhesion properties of the probiotic lactobacillus rhamnosus GG SpaCBA and SpaFED pilin subunits. *Appl. Environ. Microbiol.* 76:2049-57
295. **Vos SM, Tretter EM, Schmidt BH, Berger JM.** (2011). All tangled up: How cells direct, manage and exploit topoisomerase function. *Nat. Rev. Mol. Cell Biol.* 12:827-41
296. **Waldmann TA.** (2003). Immunotherapy: Past, present and future. *Nat. Med.* 9:269-77
297. **Wang KY, McCurdy S, Shea RG, Swaminathan S, Bolton PH.** (1993). A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry* 32:1899-904
298. **Wang Y, Rando RR.** (1995). Specific binding of aminoglycoside antibiotics to RNA. *Chem. Biol.* 2:281–90

299. **Wesolowski J, Alzogaray V, Reyelt J, Unger M, Juarez K, *et al.*** (2009). Single domain antibodies: Promising experimental and therapeutic tools in infection and immunity. *Med. Microbiol. Immunol.* 198:157-74
300. **Westra E, van Erp P, Künne T, Wong S, Staals R, *et al.*** (2012). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* 46:595-605
301. **Whitson PA, Hsieh WT, Wells RD, Matthews KS.** (1987). Influence of supercoiling and sequence context on operator DNA binding with lac repressor. *J. Biol. Chem.* 262:14592-9
302. **Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S.** (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* 1
303. **Willis MC, Collins B, Zhang T, Green LS, Sebesta DP, *et al.*** (1998). Liposome-anchored vascular endothelial growth factor aptamers. *Bioconj. Chem.* 9:573-82
304. **Wilson C, Nix J, Szostak J.** (1998). Functional requirements for specific ligand recognition by a biotin-binding rna pseudoknot. *Biochemistry* 37:14410-9
305. **Wilson C, Szostak JW.** (1995). *In vitro* evolution of a self-alkylating ribozyme. *Nature* 374:777-82
306. **Win MN, Klein JS, Smolke CD.** (2006). Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res.* 34:5670-82
307. **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, *et al.*** (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* 67:235-42
308. **Wochner A, Menger M, Orgel D, Cech B, Rimmele M, *et al.*** (2008). A DNA aptamer with high affinity and specificity for therapeutic anthracyclines. *Anal. Biochem.* 373:34-42
309. **Wojcik J, Hantschel O, Grebien F, Kaupe I, Bennett KL, *et al.*** (2010). A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat. Struct. Mol. Biol.* 17:519-27
310. **Wu AM, Senter PD.** (2005). Arming antibodies: Prospects and challenges for immunoconjugates. *Nat. Biotechnol.* 23:1137-46
311. **Wu Y, Sefah K, Liu H, Wang R, Tan W.** (2010). DNA aptamer-micelle as an efficient detection/delivery vehicle toward cancer cells. *Proc. Natl. Acad. Sci. USA* 107:5-10
312. **Xu L, Aha P, Gu K, Kuimelis RG, Kurz M, *et al.*** (2002). Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* 9:933-42
313. **Yan X, Maier CS.** (2009). Hydrogen/Deuterium Exchange Mass Spectrometry In *Mass Spectrometry of Proteins and Peptides*, ed. MS Lipton, L Paša-Tolic, 492:255-71. New York, USA: Humana Press.
314. **Ye L, Mosbach K.** (2008). Molecular imprinting: Synthetic materials as substitutes for biological antibodies and receptors. *Chem. Mat.* 20:859-68
315. **Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, *et al.*** (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J. Am. Chem. Soc.* 127:16512-7
316. **Yingfu L, Yi L, Ravelet C, Peyrin E.** (2009). Aptamers in Affinity Separations: Stationary Separation. In *Functional Nucleic Acids for Analytical Applications*:271-86: Springer New York.
317. **Yu Y, Ye L, Haupt K, Mosbach K.** (2002). Formation of a class of enzyme inhibitors (drugs), including a chiral compound, by using imprinted polymers or biomolecules as molecular-scale reaction vessels. *Angew. Chem. Int. Ed.* 41:4459-63
318. **Zafra O, Fraile S, Gutiérrez C, Haro A, Páez-Espino AD, *et al.*** (2011). Monitoring biodegradative enzymes with nanobodies raised in Camelus dromedarius with mixtures of catabolic proteins. *Environ. Microbiol.* 13:960-74
319. **Zahirul Khan IM, Opdebeeck JP, Tucker IG.** (1994). Immunopotentiation and delivery systems for antigens for single-step immunization: Recent trends and progress. *Pharm. Res.* 11:2-11
320. **Zahnd C, Amstutz P, Plückthun A.** (2007). Ribosome display: Selecting and evolving proteins *in vitro* that specifically bind to a target. *Nat. Methods* 4:269-79
321. **Zhou J, Swiderski P, Li H, Zhang J, Neff CP, *et al.*** (2009). Selection, characterization and application of new RNA HIV gp 120 aptamers for facile delivery of Dicer substrate siRNAs into HIV infected cells. *Nucleic Acids Res.* 37:3094-109
322. **Zichi D, Eaton B, Singer B, Gold L.** (2008). Proteomics and diagnostics: Let's Get Specific, again. *Curr. Opin. Chem. Biol.* 12:78-85
323. **Zourob M, Elwary S, Turner A, Warriner K, Lai EPC, *et al.*** (2008). Molecular Imprinted Polymers for Biorecognition of Bioagents. In *Principles of Bacterial Detection: Biosensors, Recognition Receptors and Microsystems*:785-814: Springer New York.
324. **Zuker M.** (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406-15

Co-author affiliations

Stan J.J. Brouns, Carel Fijen, Mark Levisson[a], John van der Oost, Javier Ramiro-Garcia, Hauke Smidt, Willem M. de Vos, Edze R. Westra.
Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703HB Wageningen, The Netherlands.

Michel H. M. Eppink.
Laboratory of Bioprocess Engineering, Wageningen University, Bomenweg 2, 6703HD Wageningen, The Netherlands

Arjan Barendregt, Esther van Duijn.
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

Bauke W. Dijkstra[b], Johan Hekelaar.
Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 7, 9747AG Groningen, The Netherlands.

Regina Stoltenburg, Beate Strehlitz.
Department of Environmental Biotechnology Centre for Environmental Research Leipzig-Halle GmbH (UFZ), Permoserstrasse 15, 04318 Leipzig, Germany.

Christophe Escudé.
CNRS UMR 7196, Muséum National d'Histoire Naturelle, INSERM U 565, Case Postale 26, 43 rue Cuvier, 75231 Paris Cedex 05, France.

Kevin Dyer, John A. Tainer.
Life Science Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, 94720 Berkeley, CA, USA.

Present address:
[a]: Wageningen UR Food & Biobased Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands.
[b]: ESRF, 6 rue Jules Horowitz, BP 220, 38043 Grenoble Cedex 09, France.

Nederlandse samenvatting

Verreweg de bekendste functie van DNA is de opslag van genetische informatie, maar DNA is veelzijdig en geschikt voor veel meer functies. DNA bestaat uit vier bouwstenen (nucleotiden) en deze vier bouwstenen vormen voor de opslag van genetische informatie twee lange ketens. Samen vormen deze ketens de voor DNA typerende dubbele helix. De vier bouwstenen kunnen in elke willekeurige volgorde achter elkaar worden gezet, maar in levende organismen is deze volgorde heel specifiek vastgelegd (sequentie). Vergelijkbaar met morse code, waar punten en streepjes coderen voor letters, coderen blokjes van drie bouwstenen voor de letters (aminozuren) waaruit eiwitten zijn opgebouwd. Aan de hand van deze code kan DNA dus worden omgezet in eiwitten (via een RNA intermediair). Eiwitten zijn verantwoordelijk voor veel belangrijke processen in levende organismen. Een foutje (mutatie) in de volgorde van de DNA bouwstenen kan ervoor zorgen dat het eiwit zijn functie (deels) verliest.

In de jaren '90 is ontdekt dat korte stukjes DNA (of RNA), van 20 tot 60 bouwstenen in een enkele streng, selectief aan bepaalde moleculen kunnen binden; zelfs als deze moleculen van nature niet binden aan DNA (of RNA). Dit soort korte DNA (of RNA) stukjes worden aptameren genoemd. De targetmoleculen kunnen zowel relatief simpele moleculen zijn, zoals sommige geneesmiddelen (diclofenac), of complexere moleculen, zoals eiwitten (trombine). Dit proefschrift gaat voornamelijk over het selecteren en karakteriseren van zulke DNA aptameren.

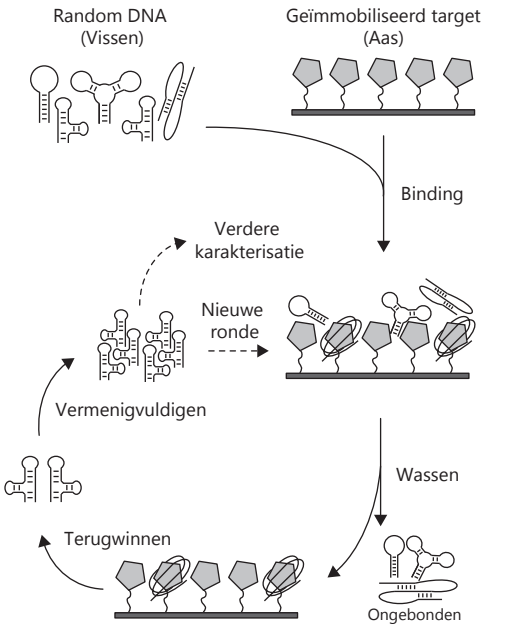
Net als al het DNA (en RNA) zijn ook aptameren opgebouwd uit vier verschillende bouwstenen (nucleotiden). De specifieke volgorde van bouwstenen in een aptameer codeert echter niet voor een eiwit, maar zorgt ervoor dat het korte stukje DNA (of RNA) een 3D-structuur vormt die het target kan binden. Van de vele miljarden mogelijke volgordes zijn er slechts enkele in staat om het target daadwerkelijk te binden, helaas is het niet mogelijk om vooraf te voorspellen welke volgorde van bouwstenen hiertoe in staat is. De uitdaging is dus om uit de vele miljarden mogelijke combinaties de juiste combinatie te vinden die het target het beste bindt.

Tegenwoordig zijn er methodes beschikbaar om miljarden korte DNA stukjes chemisch te produceren, elk met een willekeurige volgorde van de bouwstenen. De selectie van aptameren is te vergelijken met vissen in een vijver. Tijdens de selectie van een aptameer wordt het target (het aas) blootgesteld aan deze DNA stukjes (de vissen), met als gevolg dat de DNA stukjes die het target kunnen binden dit ook zullen doen. DNA stukjes die het target niet binden kunnen makkelijk worden weggewassen (deze zijn verder niet meer van belang). De gebonden DNA stukjes worden vervolgens teruggewonnen en vermenigvuldigd (zie Figuur 1). Doorgaans moeten deze stappen (blootstellen, wassen, terugwinnen en vermenigvuldigen) een aantal keer worden herhaald, om zo de meest geschikte aptameren te verrijken. Tijdens dit proces blijft de volgorde van de bouwstenen onbekend, deze kan pas in een later

stadium achterhaald worden (op enig moment heb je beet, zonder te weten hoe de vis er uit ziet; dit wordt pas duidelijk op het moment dat je de vis uit het water haalt).

De eigenschap om specifiek een bepaald molecuul te binden is echter niet exclusief voorbehouden aan aptameren, er zijn verscheidene andere groepen moleculen die hier ook toe in staat zijn. In **Hoofdstuk 1** worden deze groepen (antilichamen, bindingseiwitten en MIPs) beschreven en hun eigenschappen met elkaar vergeleken.

Een van de belangrijkste eigenschappen van aptameren is de sterkte en de specificiteit van de binding tussen het aptameer en het targetmolecuul. Over het algemeen geldt: hoe sterker en specifieker de binding hoe beter dat is. In **Hoofdstuk 2** is een geavanceerde techniek beschreven (Surface Plasmon Resonance, ofwel SPR) die inzicht kan geven in hoe snel een aptameer-target complex gevormd wordt en hoe snel het weer uit elkaar valt. Deze waarden vormen samen een maat voor hoe sterk de interactie is.



Figuur 1: Selectie van DNA aptameren. Een grote hoeveelheid DNA fragmenten, met een willekeurige samenstelling, wordt blootgesteld aan een geïmmobiliseerd target. DNA moleculen die dit target kunnen binden zullen dit doen, het overige DNA wordt weggewassen en is niet meer van belang. Vervolgens wordt het gebonden DNA teruggewonnen en vermenigvuldigd. Deze stappen worden doorgaans een aantal keer herhaald om zo de beste binder te selecteren. In een latere fase worden de geselecteerde DNA moleculen verder gekarakteriseerd.

De selectie en karakterisatie van aptameren die het eiwit streptavidine binden is beschreven in **Hoofdstuk 3**. Dit eiwit is een modelsysteem voor selectie van aptameren en is onder andere gekozen om de selectieprocedure van aptameren verder te ontwikkelen en te valideren. Uit de behaalde resultaten blijkt onder andere dat minimale verschillen in de volgorde van de bouwstenen grote invloed hebben op de sterkte van de aptameer-target interactie. Twee bouwstenen, elk op een specifieke plek, blijken zelfs essentieel; als ze vervangen worden door een andere bouwsteen is een interactie niet meer mogelijk.

In **Hoofdstuk 4** is de kristallisatie van een streptavidine-aptameer complex beschreven. Naast mooie plaatjes levert een kristalstructuur een schat aan gedetailleerde informatie op over de interactie. Helaas is het kristalliseren van eiwit-DNA complexen erg lastig gebleken, het enige kristal van voldoende kwaliteit voor

verdere analyse bleek alleen uit eiwit te bestaan.

Naast de selectie van een aptameer dat streptavidine bindt is ook het SpaC eiwit gebruikt als target voor aptameer selectie. Dit eiwit is interessant omdat het de *Lactobacillus rhamnosus* GG bacterie, een probioticum, helpt om langer in de darmen te verblijven. De productie van dit eiwit, en de aptameer selectieprocedure zijn beschreven in **Hoofdstuk 5**. Een aantal korte DNA stukjes lijken specifiek aan dit eiwit te binden maar verdere experimenten zijn nodig om dat onomstotelijk vast te kunnen stellen.

Om beter inzicht te krijgen in de processen die leiden tot de verrijking van bepaalde DNA stukjes, is de volgorde van de bouwstenen van een groot aantal DNA stukjes bepaald (>84 miljoen). Uit een voorlopige analyse van deze data, beschreven in **Hoofdstuk 6**, blijkt dat de verrijking van aptameren zeer sterk beïnvloed wordt door de wijze waarop gebonden DNA stukjes worden teruggewonnen tijdens de selectie. Met deze kennis zou de selectieprocedure voor aptameren verder geoptimaliseerd kunnen worden.

In **Hoofdstuk 7** is een methode beschreven om een circulair DNA molecuul (plasmide) te vangen op een oppervlakte geschikt voor SPR metingen, zonder daarbij de natuurlijke ordening (topologie) van het DNA te verstoren. De natuurlijke ordening van DNA is van grote invloed op de bindingssterkte van eiwitten die onder andere betrokken zijn bij de vermenigvuldiging van DNA. Bij het bestuderen van dit soort eiwit-DNA interacties is het dus van belang de natuurlijke ordening van DNA zo accuraat mogelijk na te bootsen, de beschreven methode zou daarvoor zeer geschikt kunnen zijn.

In **Hoofdstuk 8** worden de belangrijkste bevindingen uit dit proefschrift samengevat. Tevens worden nieuwe ontwikkelingen op het gebied van aptameer bouwstenen en nieuwe kansrijke toepassingen van aptameren nader beschreven.

Dankwoord

Het is gedaan. Het zit er op! Maar niet zonder dankwoord natuurlijk, want dit project zou nooit zo goed van de grond zijn gekomen zonder de bijdrage van velen en daar ben ik hen *allen* dankbaar voor. Ik kijk terug op een leuk uitdagend project, maar bovenal op een fijne tijd bij Microbiologie.

Slechts weinigen weten dat de aanzet voor dit project is terug te voeren tot Burger Bob's en Ale Works in Bozeman Montana, waar ik destijds stage liep. In een intercontinentaal telefoongesprek, tijdens diezelfde stage, werd een promotieplaats binnen BacGen en MolEco beklonken. John, deze snelle en doelgerichte gang van zaken typeert je, evenals je tomeloze enthousiasme (die is benijdenswaardig). Hauke, ten tijde van dit bewuste telefoongesprek kenden we elkaar nauwelijks, toch durfde je het avontuur aan te gaan. Dank aan beide voor jullie vertrouwen, inspiratie en adviezen, ik heb er veel van opgestoken. Aptameren hebben de ruimte helaas niet gehaald, zoals dat ooit wel de bedoeling was, maar ik waardeer het dat jullie me de ruimte hebben gegeven het project anders in te vullen toen het originele plan gaandeweg niet haalbaar bleek.

Binnen die ruimte was het mogelijk om aan SpaC te werken. Het kloneren van het gen en de productie van het eiwit gingen van een leien dakje, daarna waren de doorbraken jammer genoeg dunner gezaaid. Willem, ik vind het erg fijn dat ik aan dit project heb kunnen werken. Ik wil je bedanken voor je inspanningen om hoofdstuk 5 naar een hoger niveau te tillen, maar ook voor je bezielende leiding van de vakgroep (en de efficiënte DB vergaderingen).

Sebastian, wie kon in september 2003, aan het begin van onze studie, vermoeden dat we ooit bij dezelfde vakgroep promotieonderzoek zouden gaan doen? Door je snelle inzicht en rekenvaardigheid waren we vaak al klaar met de uitwerking van menig practicum nog voordat ik de vragen goed en wel begreep... Je kennis en hulpvaardigheid zijn ongeëvenaard, heel veel succes met het afronden van je eigen proefschrift! Mark L, in een vroeg stadium ben je bij dit project betrokken geraakt, ik denk dat we inmiddels wel van een productieve samenwerking mogen spreken. Dank voor je vele ideeën, suggesties, maar bovenal gezelligheid en goede (en minder goede) grappen. Ik vind het erg fijn en bijzonder dat jullie m'n paranimfen willen zijn!

Involvement in this project was not restricted to people within microbiology, I am grateful to all collaborators for their valuable contributions. Regina und Beate, thanks for having me in Leipzig it really was a boost for the project! I am happy we could continue our collaboration with the high throughput sequencing; to be continued... Kevin and John T, despite the difficulties we managed to extract some nice publication figures from the SAXS experiments, it worked out well. Christophe, thanks for sharing your knowledge on DNA-triplexes already at an early stage of the project. Anke and Saurabh, I am glad that our collaboration resulted in some nice papers! Esther en Arjan, bedankt voor de Native MS metingen aan de streptavidine-aptameer complexen,

we hebben de resultaten een mooi plekje kunnen geven. Bauke en Johan, helaas bleek het kristal achteraf geen DNA te bevatten, maar de resultaten hebben toch hun weg naar de literatuur gevonden. Ook dank aan Maurice Franssen, zeker in de beginfase heb ik je bedolven onder organisch chemische vragen die je altijd geduldig beantwoordde. Graag wil ik hier ook de overige leden van de promotiecommissie bedanken voor al hun inspanningen.

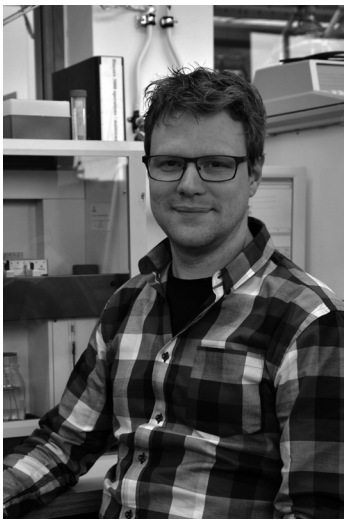
Hoogtepunten en dieptepunten kunnen dicht bij elkaar liggen, maar in beide gevallen is het fijn ze te kunnen delen met collega's, tijdens de lunch, koffie, of bij een andere gelegenheid, zoals de We-day, Veluweloop of het bierbrouwen. Een van de meest memorabele momenten was de AIO-reis langs de Amerikaanse oostkust: leerzaam, gezellig en voedzaam (...), die herinnering blijft. Herewith I would like to thank *everyone* who has spent time at Microbiology for contributing to the nice open atmosphere. Special thanks to Anja, Amos, Bart, Bas, Bram, Carel, Carolien, Chris, Clara, Daan, Detmer, Edze, Elleke, Faab, Farai, Farrakh, Gerben, Hans, Janneke, Jannie, Javi, John R, Marcel, Marco, Maria, Mark M, Matthijs, Milkha, Melvin, Naïm, Noora, Philippe, Raymond, Sjoerd, Sjon, Stan, Steven, Teunke, Tessa, Thomas, Tijn, Tom (vd B en vd W) and Wim, for your (un)conscious contributions to this thesis.

De boog kan niet altijd gespannen zijn en bij tijd en wijle is het dus goed om het onderzoek te laten voor wat het is. Ontspannen (en afzakken tot bedenkelijk niveau) lukt gegarandeerd met hulp van 'Steunpunt Westland'; dames en heren Kantelaars dank daarvoor! 'Lotgenotencontact' met oud studiegenoten is hierbij overigens ook van groot belang geweest. Hier wil ik ook Elly en Susanne bedanken voor een stukje (...) ontspanning en interesse in mijn onderzoek. Vanzelfsprekend wil ik ook mijn ouders en zusje bedanken voor steun tijdens mijn studie, stage en promotieonderzoek. Ondanks dat mijn bezigheden altijd een beetje een mysterie voor jullie zijn gebleven (hopelijk kan m'n lekenpraatje daar wat verandering in brengen...).

Onderzoek is, net als hardlopen, veel leuker als je het niet alleen hoeft te doen. Lieve Marieke, je hebt me de afgelopen jaren enorm geholpen en het beste in me naar boven gehaald. Het is fijn om mooie resultaten en frustraties te kunnen delen met iemand die in hetzelfde schuitje zit. Dank voor al je liefde en ondersteuning. Het is erg fijn om bij je te zijn, laten we dat nog lang zo houden!

Vincent

About the author



Vincent Jacobus Bernardus Ruigrok was born on the 23rd of June 1985 in 's-Gravenhage, The Netherlands. After obtaining his VWO diploma in 2003, at Het Westland College in Naaldwijk, he joined the multidisciplinary Biotechnology study-program at Wageningen University and Research Centre. During this program he developed an interest in cellular and molecular biotechnology, which led to a BSc thesis at the laboratory of microbiology at Wageningen University, to which he would return for his PhD research. During his BSc thesis he worked on the site-directed mutagenesis of a thermo-stable aldolase. This work was followed by an MSc thesis at the laboratory of molecular biology, also at Wageningen University. Here he worked on tracing gene evolution in legumes. To complete his studies, he did an internship at Montana State University, Bozeman MT, USA, at the group of prof. Mark Young. During this six months period he studied the dynamics of host-virus interactions in the extreme environments of Yellowstone National Park. In November 2008 he started his PhD-research at the laboratory of microbiology. His research focused on the selection and characterization of DNA aptamers, which can be regarded as nucleic acid analogues of antibodies. He also developed Surface Plasmon Resonance assays to determine the aptamer-target binding kinetics. Besides aptamers and SPR, he is interested in science in general but especially in biotechnological developments and the early days of molecular biology.

List of publications

Srivastava SK, **Ruigrok VJB**, Thompson NJ, Trilling AK, Heck AJR, van Rijn C, Beekwilder J, Jongsma MA. (2013). 16 kDa heat shock protein from heat-inactivated *Mycobacterium tuberculosis* is a homodimer – suitability for diagnostic applications with specific llama VHH monoclonals. *PLoS ONE* (Accepted)

Ruigrok VJB, Westra ER, Brouns SJJ, Escudé C, Smidt H, van der Oost J. (2013). A capture approach for supercoiled plasmid DNA using a triplex-forming oligonucleotide. *Nucleic Acids Research* DOI: 10.1093/nar/gkt239

Trilling AK, Harmsen MM, **Ruigrok VJB**, Zuilhof H, Beekwilder J. (2013). The effect of uniform capture molecule orientation on biosensor sensitivity: Dependence on analyte properties. *Biosensors and Bioelectronics* 40: 219-226

Ruigrok VJB [†], Levisson M [†], Hekelaar J, Smidt H, Dijkstra BW, van der Oost J. (2012). Characterization of aptamer-protein complexes by X-ray crystallography and alternative approaches. *International Journal of Molecular Sciences* 13: 10537-10552

Ruigrok VJB, van Duijn E, Barendregt A, Dyer K, Tainer JA, Stoltenburg R, Strehlitz B, Levisson M, Smidt H, van der Oost J.(2012). Kinetic and stoichiometric characterisation of streptavidin-binding aptamers. *ChemBioChem* 13:829-836

Ruigrok VJB [†], Levisson M [†], Eppink MHM, Smidt H, van der Oost J. (2011). Alternative affinity tools: more attractive than antibodies? *Biochemical Journal* 436:1-13

Brumfield SK, Ortmann AC, **Ruigrok V**, Suci P, Douglas T, Young MJ. (2009). Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus Sulfolobus turreted icosahedral virus. *Journal of Virology* 83: 5964-5970

[†] Contributed equally

Overview of completed training activities

Discipline specific activities

Meetings

- Annual meeting Protein research, Nucleic acids, and Lipids & Biomembranes. 2009. Veldhoven, NL.[1]
- NPP 3rd symposium, 2010.[2]
- Biosensors 2010. 2010. Glasgow, UK.
- DIPIA 2010. 2010. Barcelona, ES.[1]
- 100th anniversary NVvM meeting. 2011. Papendal, NL.[1,3]
- Chains congress. 2011. Maarn, NL.[1]
- 14th Netherlands biotechnology congress. 2012. Ede, Nl.[2]

[1] Poster presentation
[2] Oral presentation
[3] 3rd Poster prize

Courses

- Aptamer workshop. 2010. Centre for environmental research, Helmholtz Institute Leipzig, DE.
- Advanced biosensor workshop (level 1). 2011. Biosensor Tools, Savona, IT

General courses

- PhD competence assessment. 2009. Wageningen, NL.
- VLAG PhD week. 2009. Bilthoven, NL.
- Interpersonal communications for PhD students. 2010. Wageningen, NL.
- Techniques for writing and presenting a Scientific paper. 2010. Wageningen, NL.
- NWO-Talent Class. 2010. The Hague, NL.
- Teaching and supervising thesis students. 2011. Wageningen, NL.

Optionals

- Organic chemistry (ORC-20306). 2009. Wageningen, NL.
- PhD trip 2009, Northeast Coast, USA.
- Bacterial genetics group meetings (weekly).
- Molecular ecology group meetings (weekly).

The research described in this thesis was supported by a grant from the Netherlands Organization for Scientific Research and the Netherlands Institute for Space Research to Prof. Dr. H. Smidt. [ALW-GO-PL/08-08].

Cover illustration: An aptamer model and a titration kinetics curve (based on kinetic parameters of StrepApt3, described in chapter 3 of this thesis).
Cover design: Vincent Ruigrok
Font cover: **Myriad Pro Bold Condensed**

Typeset: InDesign® CS6
Font inner works: Segoe UI

Printed by: Gildeprint Drukkerijen - The Netherlands

Printing of this thesis was financially supported by:
Eurogentec, Seraing, Belgique